REMARKS

Claims 109-115, 117, 119-133, and 135-166 are currently pending. Claims 112-115, 117, and 128-130 are withdrawn. Claims 1-108, 116, 118, and 134 are cancelled. Claims 109, 127, and 131 are currently amended. Claims 143-166 are newly added. The new claims do not add new matter. Support for the claims may be found, in part, in claims 112-118, Figure 24, and paragraph [0056] of the published application.

I. Summary of telephone conference

Applicants appreciate the April 19th phone conference with Examiner Bhat. During the phone conference, it was determined that *one way* of addressing the §103 rejections of record is to define R1 as an antibody, as Egholm et al. and Weston et al. do not disclose antibody epitope binding agents. Even though there are other differences between the cited art and the presently pending claims, to expedite prosecution, Applicants have so amended claim 109.

Additionally, Examiner Bhat expressed concern over the free energy limitation of claim 109. Specifically, the Examiner asserted that the free energy limitation was (A) broad, and (B) does not identify a specific structural component for its calculation. Applicants disagree with both points, as detailed below.

(a) the free energy limitation is not overly broad

Claim 109 provides the range of acceptable free energy for the duplex comprising R3 and R7, i.e. about 5.5 kcal/mol to 8.0 kcal/mol. As stated in earlier Applicant responses, this free energy range is essential to maintaining the internal balance of the claimed biosensor, between a false-positive signal on the one hand, and lack of sensitivity on the other. Hence, this range is not broad or arbitrary, but rather, reflects the narrow range of free energy values across which the claimed biosensor will function. Stated another way, if R3 and R7 of a biosensor do not meet the free energy requirements of claim 109, the biosensor will not produce a detectable signal (i.e., the biosensor will not work).

To clarify this requirement, claim 109 has been amended to further reflect this internal balance. In particular, claim 109 states that "R3 and R7 are a pair of complementary nucleotide sequences having a free energy for association … **such that R3 and R7 only associate when R1 and R5 are bound to the target molecule**." This amendment explicitly incorporates the internal balance, discussed above, into the claim limitations.

Additionally, claim 109 provides a researcher skilled in the art with the necessary parameters to calculate the free energy for R3 and R7. In the context of nucleic acid duplex formation, free energy depends on three parameters: temperature, salt concentration, and the nucleic acid sequence of the duplex. (See the § 1.1.32 declaration of Dr. Tomasz Heyduk: "For a given nucleic acid duplex, the free energy depends on three factors…") Claim 109 provides a skilled researcher with two of these three parameters, i.e. temperature and salt concentration.[1] As a result, a researcher would be able to determine, for a given sequence (i.e. the third parameter), if the sequence will have the necessary free energy within the applicable temperature and salt ranges of claim 109.[2]

The Office has expressed concern that variations in methods of determining free energy make it difficult to determine whether a given sequence would fall within R3 and R7 of claim 109. This concern is based, in part, on the Office's reliance on the PNAS paper by SantaLucia. *Applicants assert that this reliance has misled the Examiner*. Scientifically speaking, there is only one definitive way to determine free energy, and that is to actually perform the experiment. To circumvent performing this experiment, however, researchers have attempted to create a system for *estimating* free energy. These attempts have been summarized, in part, in the Santa Lucia PNAS paper, which is directed to the nearest neighbor method of estimating free energy.

The nearest neighbor theory is based on the premise that the free energy of a given complementary base pair is contingent on the context of that base pair

---

[1] Heyduk declaration, at point 3a.
[2] Id., at point 3c.

in the surrounding nucleic acid sequence, i.e., the neighbors.[3] From this basic premise, different groups used different means to come up with a system for estimating free energy. Hence, the different sets of data summarized in Table 1 of the SantaLucia PNAS paper do **NOT** represent different **methods** of experimentally calculating free energy, but rather, represent different ***estimations*** of nearest neighbor parameters. Stated another way, **the ambiguity the Office was attributing to the METHOD of calculating free energy is instead properly attributed to ESTIMATIONS of nearest neighbor parameters**. For instance, the abstract of the SantaLucia PNAS paper explains that the different values in Table 1 come from different methods of determining the estimations:

> The seven studies [represented in Table 1] used data from natural polymers, synthetic polymers, oligonucleotide dumbbells, and oligonucleotide duplexes to derive [nearest neighbor] parameters; used different methods of data analysis; used different salt concentrations; and presented the [nearest neighbor] thermodynamics in different formats.

Hence, the different values in Table 1 are solely attributed to different methods of estimating free energy, as opposed to different methods of experimentally determining free energy. In fact, scientifically speaking, there is only one definitive way to determine free energy, and that is to actually perform the experiment. All other methods are designed to provide "estimations," not actual free energy values. As a result, Applicants assert on the record that the free energy for a given nucleotide sequence should be directly determined, via experiment, for the purposes of ascertaining whether the sequence falls within the range of claim 109. This is consistent with the state of the art, which is evidenced in part by the § 1.132 declaration by Dr. Tomasz Heyduk submitted herewith. Furthermore, by making this statement on the record, Applicants are bound by the theory of prosecution history estoppel. Consequently, in any future

---

[3] Id., at point 4.

disputes over claim construction, Applicants will be held to experimentally determining free energy.

(b) The free energy requirement is tied to a structural component of the biosensor.

The Office has asserted that the free energy limitation of claim 109 does not identify a specific structural component. Applicants disagree.

The free energy limitation of claim 109 is specifically tied to the composition of R3 and R7 and meaningfully limits the composition of R3 and R7. As stated above, if the free energy for R3 and R7 falls outside the free energy range of claim 109, the biosensor will not work. In essence, the free energy limitation on R3 and R7 goes to the heart of the invention.

Furthermore, one of skill in the art, given the temperature and salt conditions highlighted in claim 109, would be able to determine whether a given sequence for R3 and R7 would possesses a free energy within the claimed range, by experimentally determining the free energy as described above.

II. Declarations previous filed

Applicants would like to clarify a statement made by the Office in regard to the Heyduk declaration filed on September 11, 2009. Specifically, the Office states at page 3 of the Office Action mailed 12/18/2009 that: "it is not clear which of these components contribute or inhibit or control the unexpected signal production or the R3/R7 pair having free energy of association between about 5.5 kcal/mole and 8.0 kcal/mole..."

As stated in the response filed September 11, 2009, on pages 18 and 19, the ***only*** difference between the two biosensors used in the declaration experiment was the sequence of R3/R7. Stated another way, the other parameters listed by the Office (i.e. antibodies, Tris buffer, EDTA, and BSA) were ***identical between the two sensors***, and therefore, ***could not have contributed to the unexpected results***. The experiment was designed in this

manner to illustrate the importance of the free energy range to the functioning of a biosensor of claim 109. The Baez sequence, which had a free energy outside of the claimed range, did not produce a functional biosensor, as illustrated by the diagram in the September 11th declaration.

III. Claim interpretation, 35 USC 112, sixth paragraph

Applicants agree with the Office's interpretation of "detection means" in claims 109, 127, and 131.

IV. § 103 rejection in view of Egholm et al. and Weston et al.

Reconsideration is requested of the rejection of claims 109-110, 119-127, 131-132, and 135-142 under 35 USC § 103(a) in view of Egholm et al. and Weston et al.

Amended claim 109 is directed to a biosensor comprising two constructs (R1-R2-R3-R4 and R5-R6-R7-R8). Claim 109 requires that R1 and R5 are antibodies. The Office has previously acknowledged that neither Egholm nor Weston disclose a reporter where R1 or R5 are antibodies. As a result, Applicants request withdrawal of the rejection in light of Egholm and Weston.

Specifically, Egholm et al. discloses a reporter that recognizes and detects nucleic acid targets. The Egholm reporter ***only*** uses nucleic acid as an epitope binding agent, because the Egholm reporter relies on hybridization between the nucleic acid epitope binding agent and the target to produce a signal. Nowhere does Egholm teach or suggest an antibody based epitope binding agent, as required by claim 109.

Similarly, Weston et al. discloses a reporter that recognizes and detects ***only*** nucleic acid targets. The Weston reporter only uses nucleic acid as an epitope binding agent, because the Weston reporter relies on hybridization between the nucleic acid target and the nucleic acid epitope binding agent to produce a signal. Nowhere does Weston teach or suggest an antibody based epitope binding agent as required by claim 109.

Three criteria must be present to establish a *prima facie* case of obviousness.[4] First, the prior art reference must teach or suggest *all* the claim limitations. Second, there must be some suggestion or motivation in the knowledge generally available to one of ordinary skill in the art to modify the reference. Third, there must be a reasonable expectation of success.[5] Not one of these three criteria is satisfied by the combination of Egholm et al. and Weston et al. cited by the Office.

(a) the cited art does not disclose antibody epitope binding agents

As discussed above, neither Egholm nor Weston disclose R1 and R5 of the present claims. Specifically, neither Egholm nor Weston disclose a biosensor where R1 and R5 are antibodies. Hence, Egholm and Weston, whether taken individually or combined, fail to disclose each element of the pending independent claims.

(b) no motivation or reasonable expectation of success

As neither Egholm nor Weston disclose a reporter that comprises R1 and R5 of the present claims, there is no motivation to combine them to create a sensor where R1 and R5 are antibodies, nor is there any reasonable expectation of success in doing so. Both the Egholm and the Weston reporter recognize their target via nucleic acid hybridization, necessitating that the epitope binding agent in both references is a nucleic acid. Hence, there is no motivation in the cited art to combine Egholm and Weston to arrive at the biosensor of claim 109.

(c) dependent claims

Claims 110 and 119-126 depend from claim 109 and necessarily incorporate each limitation of claim 109. As a result, Egholm and Weston do not

---

[4] MPEP §2143
[5] *Id.*

render claims 110, and 119-126 obvious for the same reasons as cited above with respect to claim 109.

Claims 132 and 135-142 depend from claim 131 and necessarily incorporate each limitation of claim 131. Specifically, claims 132 and 135-142 each incorporate the limitation that R1 and R5 are antibodies. Egholm and Weston only disclose nucleic acid epitope binding agents. As a result, Egholm and Weston do not render claims 132 and 135-142 obvious for the same reasons as cited above with respect to claim 132.

Consequently, Applicants respectfully request withdrawal of the rejection of claims 109-110, 119-127, 131-132, and 135-142 as obvious in light of Egholm and Weston. As this is the only remaining rejection for claims 110, 119-127, 132, and 135-142, Applicants believe these claims to be in condition for allowance.

(d) new claims

New claims 143-154 encompass a biosensor where R1 and R5 recognize a non-nucleic acid target molecule. As detailed above, Egholm et al. and Weston et al. only teach nucleic acid targets. As a result, Egholm et al. and Weston et al. do not render new claims 143 –154 obvious.

New claims 155-166 encompass a biosensor where R1 is an antibody epitope binding agent and R5 is an epitope binding agent. As detailed above, Egholm et al. and Weston et al. only teach nucleic acid epitope binding agents. As a result, Egholm et al. and Weston et al. do not render new claims 155 –166 obvious.

V. §103 rejection in view of Egholm et al. and Weston et al. and Baez et al.

Reconsideration is requested of the rejection of claims 109, 111, 116, 131, 133, and 134 under 35 USC §103(a) in view of Egholm et al., Weston et al., and Baez et al.

As stated above, Egholm et al. and Weston et al. do not disclose antibody epitope binding agents. To address this shortcoming, the Office has resorted to

combining Egholm et al. and Weston et al. with the Baez application. Resort to Baez, however, does not cure the defect in the Office's obviousness rejection. Specifically, as detailed below, Egholm et al. does not disclose an R3 that meets the free energy requirements of claim 109, and as acknowledged by the Office, Baez does not disclose an equivalent to the R2-R3 portion of claim 109. Hence, even in combination, Egholm et al., Weston et al, and Baez et al. do not disclose each limitation of claim 109.

(a) there is no prima facie case of obviousness

Three criteria must be present to establish a *prima facie* case of obviousness.[6] First, the prior art reference must teach or suggest ***all*** the claim limitations. Second, there must be some suggestion or motivation in the knowledge generally available to one of ordinary skill in the art to modify the reference. Third, there must be a reasonable expectation of success.[7] Not one of these three criteria is satisfied by the combination of Egholm et al., Weston et al., and Baez et al. as cited by the Office.

i. the cited references DO NOT disclose all limitations of claims 109 and 131

As summarized above, and detailed in the response filed September 11, 2009, the Baez application fails to disclose an equivalent to R3.

While Egholm et al. discloses a reporter that can detect a nucleic acid sequence, Egholm et al. **DOES NOT** disclose R3 of claim 109 as asserted by the Office. Specifically, Egholm et al. **DOES NOT** disclose a complementary nucleotide sequence having a free energy for association over the entire length of the nucleotide sequence, from about 5.5 kcal/mole to 8.0 kcal/mole. Rather, Egholm describes a complementary nucleic acid sequence that comprises ***peptide nucleic acid (PNA)***. The Office has mistakenly assumed that the sequence 'TGACTGAC' disclosed in Figure 7 of the Egholm application is DNA.

---

[6] MPEP §2143
[7] *Id.*

IT IS NOT. Instead, the sequence is PNA, as detailed in Figure 11, and Example 2 (lines 10-11). As a result, the free energy estimated by the Office ***IS NOT CORRECT***, because the free energy of a PNA/PNA duplex ***IS NOT THE SAME*** as the free energy of a DNA/DNA duplex. This is illustrated, in part, by the data in Table 1 of the attached Sen et al. reference.

Table 1 of Sen et al. shows that a PNA/PNA duplex ***DOES NOT HAVE*** the same free energy as a DNA/DNA duplex with the same sequence. In particular, the PNA/PNA duplex has a free energy that is ~2x the free energy of the identical DNA/DNA duplex in the absence of an organic solvent. The Office has cited to SantaLucia et al. as support for the free energy value erroneously assigned to Egholm et al., but SantaLucia is **ONLY APPLICABLE TO DNA**, not PNA. As a result, the Office has not established that Egholm et al. discloses a complementary sequence with a free energy from about 5.5 kcal/mole to 8.0 kcal/mole.

Hence, Egholm et al. and Weston et al., whether taken by themselves, or in combination with Baez et al., do not disclose each limitation of claim 109.

ii. no motivation or reasonable expectation of success

Not only do the references not disclose each limitation of claim 109, but the cited references provide no motivation to modify the references to arrive at the biosensor of claim 109. Specifically, none of the cited references disclose or suggest the importance of the free energy range restriction on R3. As a result, the references provide no motivation to modify any of the cited references to arrive at the required R2-R3 of claim 109.

The Office has asserted that a skilled artisan would be motivated to "modify the epitope binding agent of Egholm et al. with the expected benefit of improving the ratio of analyte specific signal to analyte non-specific background signal thereby enhancing analyte detection at very low concentration as taught by Baez et al." This assertion, however, is incorrect for at least two reasons. First, ***even if*** Baez is used to modify the epitope binding agents of Egholm, the

Office still has not established a prima facie case of obviousness, because neither Egholm nor Baez disclose R2-R3. Second, the Office is quoting Baez out of context. Paragraph [0017], read in context with the previous paragraphs, does not call out the antibody epitope binding agent as a means of improving signal. Rather, it calls out the Baez method of signal amplification as a means to improve signal. This is apparent from reading paragraphs [0014] - [0017]:

> [N]on-specific binding-signal due to non-selective binding of reporter conjugates to walls of the reaction tubes or to solid-phase reagents used in the assays even in the absence of analyte, is a serious problem in **immunoassays**.... Despite numerous attempts in this art ... non-specific binding has not been eliminated.

Hence, the improved signal referenced in paragraph [0017] of Baez is not because of the use of antibodies, but due to the claim that the Baez reporter eliminates non-specific signal derived from reporter conjugates binding to the solid surfaces of the assay. As a result, there is no motivation to combine the references to arrive at the biosensor of claim 109.

Furthermore, there is no reasonable expectation of combining the references to arrive at a biosensor of claim 109. Egholm et al. is directed solely to detecting nucleic acid targets. Generally speaking, antibodies do not recognize nucleic acid targets with the same affinity or specificity as a complementary sequence. As a result, a skilled artisan would not predict that the hybridization schemes illustrated and taught by Egholm would work if the nucleic acid epitope binding agent was replaced with an antibody of Baez.

In conclusion, the Office has not set forth a prima facie case of obviousness because the cited references do not disclose each element of claim 109, provide no motivation for modifying the cited references to arrive at claim 109, and there is no reasonable expectation of success in modifying the references. Consequently, Applicants respectfully request withdrawal of the

rejection of claim 109 as obvious in view of the combination of Egholm et al., Weston et al., and Baez et al.

(b) claims 110, 119-126, 131, 132, 135-142

Claims 110 and 119-126 depend from claim 109 and necessarily incorporate each limitation of claim 109. As a result, Egholm, Weston, and Baez do not render claims 110, and 119-126 obvious for the same reasons as cited above with respect to claim 109.

Claims 132 and 135-142 depend from claim 131 and necessarily incorporate each limitation of claim 131. Claim 131, and as a result claims 132 and 135-142, each incorporate the limitations on R1, R2, R3, R4, R5, R6, R7, and R8 found in claim 109. As detailed above, the cited combination of art does not disclose R2-R3 of claim 109. As a result, Egholm, Weston, and Baez do not render claims 131, 132 and 135-142 obvious for the same reasons as cited above with respect to claim 109.

Consequently, Applicants respectfully request withdrawal of the rejection of claims 109-110, 119-127, 131-132, and 135-142 as obvious in light of the combination of Egholm et al., Weston et al. and Baez.

(c) new claims

New claims 143-154 encompass a biosensor where R1 and R5 recognize a non-nucleic acid target molecule. R2 and R3 of the new claims are identical to claim 109. As detailed above, the combination of Egholm et al.,Weston et al., and Baez et al. do not teach R2-R3. As a result, the cited art does not render new claims 143 –154 obvious.

New claims 155-166 encompass a biosensor where R1 is an antibody epitope binding agent and R5 is an epitope binding agent. As detailed above, Egholm et al. and Weston et al. only teach nucleic acid epitope binding agents, and the combination of Egholm et al.,Weston et al., and Baez et al. do not teach R2-R3. As a result, the cited art does not render new claims 155 –166 obvious.

CONCLUSION

In light of the foregoing, applicants request entry of the claim amendments, withdrawal of the claim rejections, and solicit an allowance of the claims. The Examiner is invited to contact the undersigned attorney should any issues remain unresolved.

Respectfully submitted,

Polsinelli Shughart PC

Date: June 18, 2010

By: /Rebecca C. Riley-Vargas/
Rebecca C. Riley-Vargas, Reg. No. 60,046
110 South Fourth Street, Suite 1100
St. Louis, MO 63102
Phone – 314-889-8000
Fax – 314-231-1776
Attorney

RCR/sat

Improved Nearest-Neighbor Parameters for Predicting DNA Duplex Stability†

John SantaLucia, Jr.,* Hatim T. Allawi, and P. Ananda Seneviratne

Department of Chemistry, Wayne State University, Detroit, Michigan 48202

Received August 14, 1995; Revised Manuscript Received December 1, 1995®

ABSTRACT: Thermodynamic data were determined from UV absorbance vs temperature profiles of 23 oligonucleotides. These data were combined with data from the literature for 21 sequences to derive improved parameters for the 10 Watson−Crick nearest neighbors. The observed trend in nearest-neighbor stabilities at 37 °C is GC > CG > GG > GA ≈ GT ≈ CA > CT > AA > AT > TA (where only the top strand is shown for each nearest neighbor). This trend suggests that both sequence and base composition are important determinants of DNA duplex stability. On average, the improved parameters predict $\Delta G°_{37}$, $\Delta H°$, $\Delta S°$, and T_M within 4%, 7%, 8%, and 2 °C, respectively. The parameters are optimized for the prediction of oligonucleotides dissolved in 1 M NaCl.

Accurate prediction of DNA thermal denaturation is important for several molecular biological techniques including PCR[1] (Saiki et al., 1988), sequencing by hybridization (Fodor et al., 1993), antigene targeting (Freier, 1993), and Southern blotting (Southern, 1975). In these techniques, choice of a nonoptimal sequence or temperature can lead to amplification or detection of wrong sequences (Steger, 1994). Furthermore, knowledge of the sequence dependence of DNA melting is important for understanding the details of DNA replication, mutation, repair, and transcription (Mendelman et al., 1989; Petruska et al., 1988).

One widely used method for predicting nucleic acid duplex stability, pioneered by Tinoco and co-workers (Borer et al., 1974), uses a nearest-neighbor model for helix propagation. Several nearest-neighbor parameter sets for predicting DNA duplex stability are available in the literature (Gotoh & Tagashira, 1981; Ornstein & Fresco, 1983; Vologodskii et al., 1984; Wartell & Benight, 1985; Breslauer et al., 1986; Aida, 1988; Otto, 1989; Quartin & Wetmur, 1989; Klump, 1990; Delcourt & Blake, 1991; Doktycz et al., 1992). The quantum mechanical studies (Ornstein & Fresco, 1983; Aida, 1988; Otto, 1989) were performed in the gas phase and neglected solvent and counterion interactions and, thus, do not reflect the conditions typically found *in vivo* or *in vitro*. Data for the thermal denaturation of polymers are difficult to interpret properly since their transitions are typically not two-state (i.e., many unfolding intermediates are possible) and their melting temperatures are high. Thus, polymer melting is typically performed in solutions with low salt concentration and the thermodynamic results are extrapolated to the standard state temperature (25 or 37 °C) and higher salt concentration (Breslauer et al., 1986). In addition, polymer melting does not involve a bimolecular initiation event and is dependent on only eight invariants which are linear combinations of the 10 nearest neighbor parameters required to predict oligonucleotide thermodynamics (Gray & Tinoco, 1970; Vologodskii et al., 1984; Doktycz et al., 1992). Thus, studies of polymer thermodynamics (Gotoh & Tagashira, 1981; Vologodskii et al., 1984; Wartell & Benight, 1985; Klump, 1990; Delcourt & Blake, 1991) are most applicable for the prediction of polymer behavior but do not reliably predict oligonucleotide thermodynamics (Sugimoto et al., 1994). Thus, we decided to expand the DNA oligonucleotide thermodynamic database and derive new nearest-neighbor parameters in 1 M NaCl buffer.

In this paper, thermodynamic measurements are reported for 26 oligonucleotides ranging in length from 4 to 16 base pairs. Thermodynamic data for 23 of these sequences are combined with data for 21 oligonucleotides from the literature to derive improved nearest-neighbor parameters. The parameters are able to predict the stabilities of DNA duplexes within the limits of the nearest-neighbor model.

MATERIALS AND METHODS

DNA Synthesis and Purification. Oligonucleotides were the gift of Hitachi Chemical Research and were synthesized on solid support using standard phosphoramidite chemistry (Brown & Brown, 1991). Oligomers were removed from solid support and base blocking groups were removed by treatment with concentrated ammonia at 50 °C overnight. Each sample was evaporated to dryness, dissolved in 250 μL of water and purified on a Si500F thin-layer chromatography plate (Baker) by eluting for 5 h with *n*-propanol/ammonia/water (55:35:10 by volume) (Chou et al., 1989). Bands were visualized with an ultraviolet lamp, and the least mobile band was cut out and eluted three times with 3 mL of distilled deionized water. The sample was then evaporated to dryness. Oligonucleotides were desalted and further purified with a Sep-pak C-18 cartridge (Waters). The DNA was eluted with 30% acetonitrile buffered with 10 mM ammonium bicarbonate at pH 7.0. The purity of oligonucleotides was checked by analytical C-8 HPLC (Perceptive Biosystems) and was greater than 95%.

Measurement of Melting Curves. Absorbance vs temperature profiles (melting curves) were measured with an Aviv 14DS UV−vis spectrophotometer with a five cuvette thermoelectric controller. Custom-manufactured microcuvettes (Hellma Cells) with 0.1, 0.2, 0.5, and 1.0 cm path lengths (60, 120, 300, and 600 μL volumes, respectively) were used

† This work was supported by Hitachi Chemical Research.

* Author to whom correspondence should be addressed.

® Abstract published in *Advance ACS Abstracts,* February 15, 1996.

[1] Abbreviations: Na_2EDTA, disodium ethylenediaminetetraacetate; eu, entropy units (cal/K mol); HPLC, high-performance liquid chromatography; T_M, melting temperature; PCR, polymerase chain reaction; UV, ultraviolet.

0006-2960/96/0435-3555$12.00/0

so that melting curves could be measured with high sensitivity over a 100-fold range in oligonucleotide concentration. Aluminum adapters were used to properly position microcuvettes in the light beam and provide optimal thermal contact with the thermoelectric cuvette holder. To prevent water condensation at low temperatures, the sample compartment was purged with dry nitrogen gas.

The temperature was monitored with the temperature transducer (Analog Devices Inc.) mounted in the spindle of the Aviv thermoelectric cuvette holder. The temperature readings from the transducer were calibrated by measuring the voltage produced by a type K thermocouple inserted in a 1 cm microcuvette during a typical thermal denaturation run. We estimate the temperature measurement to be reproducible within 0.1 °C and accurate within 0.3 °C.

Oligonucleotides were dissolved in 1.0 M NaCl, 10 mM sodium cacodylate, and 0.5 mM Na_2EDTA, pH 7, buffer. Samples were "annealed" and degassed by raising the temperature to 85 °C for 5 min and then cooling to −1.6 °C over a period of 25 min just prior to a melting experiment. While at 85 °C, the absorbances were measured at 260 nm for later calculation of oligonucleotide concentration using extinction coefficients calculated from dinucleoside monophosphates and nucleotides, as described (Richards, 1975). Oligonucleotide concentration was varied over an 80−100-fold range. Samples were then heated at a constant rate of 0.8 °C/min with data collection beginning at 0 °C and ending at 90−95 °C. Control experiments with a heating rate of 0.4 °C/min gave same results as those obtained with 0.8 °C/min indicating that thermal equilibrium was established. The duplex to coil transition was monitored by measuring the absorbance at 280 nm. Air was used for the reference light beam. The spectral bandwidth was 1 nm. Absorbances were not corrected for thermal expansion since the correction was linear and less than 3% from 0 to 90 °C.

Determination of Thermodynamic Parameters. Thermodynamic parameters for duplex formation were obtained from melting curve data using the program MELTWIN v2.1 (Jeff McDowell and Douglas H. Turner, unpublished) as described (Petersheim & Turner, 1983). Data were truncated so that the upper and lower temperature baselines reflected the slopes in the transition region, generally using $T_M \pm 30$ °C (Petersheim & Turner, 1983). The root mean square difference between data and calculated curves is less than 0.5%, the approximate error in the absorbance reading. The enthalpy and entropy for the random coil to duplex equilibrium were obtained by two methods: (1) $\Delta H°$ and $\Delta S°$ from the fits of individual curves were averaged, and (2) plots of reciprocal melting temperature (T_M^{-1}) versus the natural logarithm of the total strand concentration (ln C_T) were fit to eq 1 (Borer et al., 1974):

$$T_M^{-1} = R/\Delta H° \ln C_T + \Delta S°/\Delta H° \quad (1)$$

The T_M is defined as the temperature at which half of the strands are in the double helical state and half are in the "random coil" state. For self-complementary oligonucleotides, the T_M for individual melting curves was calculated from the fitted parameters using

$$T_M = \Delta H°/(\Delta S° + R \ln C_T) \quad (2)$$

where R is the gas constant [1.987 cal/(K mol)], and the T_M is given in K. For non-self-complementary molecules, C_T in eqs 1 and 2 was replaced by $C_T/4$. Both methods are essentially a van't Hoff analysis of the data, assuming the transition equilibrium involves only two states (i.e., duplex and random coil). We also assume that the difference in heat capacities (ΔC_p) of these states is zero (Petersheim & Turner, 1983; Freier et al., 1986b). These two methods depend differently on the two-state approximation. For a given oligonucleotide, agreement of parameters derived by the two methods is a necessary, but not sufficient, criterion to establish the validity of the two-state approximation (SantaLucia et al., 1990; Marky & Breslauer, 1987). The methods described above have been shown to give thermodynamic results in good agreement with those obtained by calorimetry (Albergo et al., 1981).

Choice of Sequences. Sequences were designed or selected from the literature to meet the following criteria: (1) two-state thermodynamics, (2) T_M between 20−65 °C to minimize extrapolation to 37 °C and allow the upper and lower temperature baselines to be adequately defined, (3) sequences with three or more consecutive guanine residues are not included since these sequences consistently yield lower than expected $\Delta H°$ values (Breslauer et al., 1986), (4) sequences have terminal GC base pairs to minimize helix "fraying" that could invalidate the two-state approximation, and (5) the oligonucleotides were dissolved in 1 M NaCl so that length-dependent counterion condensation effects could be neglected (Record & Lohman, 1978; Olmsted et al., 1989). Sequences measured in this study were designed to meet the above criteria and provide uniform representation of the 10 different nearest neighbors in the database. Throughout this paper nearest-neighbor base pairs are represented with a slash separating strands in antiparallel orientation (e.g., AC/TG means 5′-AC-3′ Watson−Crick base paired with 3′-TG-5′). The 10 nearest-neighbor sequences occur in this study with the following frequencies AA/TT = 43, AT/TA = 21, TA/AT =14, CA/GT = 28, GT/CA = 27, CA/GT = 23, GA/CT = 36, CG/GC = 35, GC/CG = 33, and GG/CC = 29.

Determination of Nearest-Neighbor Thermodynamic Parameters. The total difference in the free energy of the folded and unfolded states of a DNA duplex can be approximated with a nearest-neighbor model:

$$\Delta G_i(\text{total}) = \sum_j n_{ij} \Delta G_j + \Delta G(\text{init}) + \Delta G_i(\text{sym}) \quad (4)$$

where each different oligonucleotide duplex is given the subscript i, ΔG_j are the free energies for the 10 possible Watson−Crick nearest-neighbor stacking interactions [e.g., $\Delta G_1 = \Delta G°_{37}$ (AA/TT), $\Delta G_2 = \Delta G°_{37}$ (TA/AT), ..., etc.], n_{ij} is the number of occurrences of each nearest neighbor, j, in each sequence, i, $\Delta G(\text{init})$ is the initiation free energy, and $\Delta G_i(\text{sym})$ equals +0.4 kcal/mol if duplex i is self-complementary and zero if it is non-self-complementary (Bailey & Monahan, 1978; Cantor & Schimmel, 1980).

The thermodynamic results from 44 sequences with two-state transitions were used to construct appropriate matrices for input into the linear regression analysis. The $[\Delta G_i(\text{total}) - \Delta G_i(\text{sym})]$ formed the column matrix G_{Tot} and the standard errors in the $\Delta G_i(\text{total})$, σ_i, formed the column matrix σ. The number of occurrences of each of the nearest neighbors, along with the initiation parameter formed the "stacking matrix", S with dimensions 44 × 11. The values of the 10 nearest neighbors and initiation, G_j, are unknown

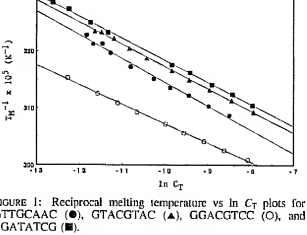

FIGURE 1: Reciprocal melting temperature vs ln C_T plots for GTTGCAAC (●), GTACGTAC (▲), GGACGTCC (○), and CGATATCG (■).

and form the column matrix, $\mathbf{G}_{NN}$. The data for all sequences is thus written:

$$\mathbf{G}_{Tot} = \mathbf{S}\mathbf{G}_{NN} \tag{5}$$

The solution of eq 5 for the nearest neighbors, $\mathbf{G}_{NN}$, was obtained using singular value decomposition (Press et al., 1989) which effectively minimizes the error weighted squares of the residuals (Bevington, 1969):

$$\chi^2 = \sum_{ij} |(\Delta G_i - S_{ij}\Delta G_j)/\sigma_i|^2 \tag{6}$$

Analogous calculations were performed to obtain nearest-neighbor parameters for ΔH° and ΔS°. All matrix manipulations were performed using the program MATHEMATICA (Wolfram, 1992). To verify our calculation methods, we derived the nearest-neighbor parameters for RNA and reproduced the literature values (Freier et al., 1986a).

RESULTS

Thermodynamic Data. All sequences in this study displayed monophasic melting transitions (data not shown) and showed concentration-dependent T_Ms, indicating complexes with molecularity greater than 1. Plots of T_M^{-1} versus ln C_T were linear (correlation coefficient > 0.99) over the entire 80−100-fold range in concentration and are shown in Figure 1 and Supporting Information (see paragraph at the end of the paper). Thermodynamic parameters derived from the average of fits of individual melting curves and from T_M^{-1} versus ln C_T are listed in Table 1. Those sequences in which the ΔH° from the two methods agree within 20% are listed in Table 1 as two-state transitions (SantaLucia et al., 1990; Marky & Breslauer, 1987). Those with differences in ΔH° greater than 20% are listed as non-two-state transitions. Experimental heat capacity differences (Table 1), ΔC_p, were determined from the slope of ΔH° vs T_M plots (data not shown), where the ΔH°s and T_Ms are from the fitted curves of each oligonucleotide at different concentrations (Petersheim & Turner, 1983; Freier et al., 1986b). For two-state transitions, parameters derived from the average of the fits and from T_M^{-1} vs ln C_T plots are equally reliable; thus the average of these parameters (Table 2) was used for the linear regression to determine nearest-neighbor increments (Table 3).

Nearest-Neighbor Parameters. Table 3 lists the nearest-neighbor thermodynamic parameters derived by multiple linear regression from the data in Table 2. These parameters allow for self-consistent and accurate prediction of the 44 sequences in Table 2 with two-state thermodynamics. Sequences with terminal T-A base pairs were not included in the regression analysis in order to minimize systematic errors due to terminal "fraying" in these sequences. We performed a control experiment where the eight sequences in Table 2 with terminal A-T base pairs were included in the regression analysis. The results showed that a poorer fit (as judged by the χ^2 and Q parameters) was observed. This was particularly true for the ΔH° and ΔS° parameters. Inclusion of these sequences systematically made nearest neighbors with 3′-terminal T residues more stable than those with 5′-terminal T residues. For example, the ΔG°_{378} for AT/TA, TA/AT, GT/CA, TG/AC, CT/GA, and GT/CA nearest neighbors were −1.48, −0.11, −1.76, −0.99, −1.53, and −1.10 kcal/mol, respectively, while the parameters for AA/TT, CG/GC, GC/CG, GG/CC, and initiation did not change compared to the values listed in Table 3. We find that the parameters in Table 3 can predict these sequences reasonably well if a penalty of +0.4 kcal/mol is assigned (for ΔG°_{37} and ΔH°) for each terminal 5′-T·T-A-3′ base pair. Note that sequences with terminal 5′-A·T-3′ base pairs are not assigned this penalty. Apparently, sequences with terminal 5′-T·A-3′ base pairs "fray" more than sequences with terminal 5′-A·T-3′ base pairs. The parameter for terminal 5′-T·A-3′ base pairs is included in Table 3.

The Helix Initiation Parameter. The ΔG°_{37} for helix initiation is +1.82 ± 0.24 kcal/mol (Table 3). This number applies to DNA duplexes with at least one G-C base pair and agrees reasonably well with the value of +2.3 kcal/mol determined previously (Pohl, 1974; Turner et al., 1990). Previous work indicated that initiation in sequences with only A-T base pairs is +3.4 kcal/mol (Scheffler et al., 1970; Turner et al., 1990). Thus, we have assumed the initiation at A-T pairs is +2.8 ± 1 kcal/mol (Table 3). This allows for the correct prediction of the ΔG°_{37} for A_8/T_8 (Table 2; Sugimoto et al., 1991) and TTTTATAATAAA/AAAATAT-TATTT (Bolewska et al., 1984).

The penalty for duplex initiation was assumed to be purely entropic for the reasons described previously (Freier et al., 1986a). When the initiation enthalpy was allowed to vary in the regression analysis, a favorable value with a large error was observed (−7.2 ± 3.4 kcal/mol), and the enthalpy increments for the 10 nearest-neighbors are less favorable by 1.1 kcal/mol, on average. When the initiation entropy was allowed to vary in the regression analysis, a more unfavorable value was observed (−29.1 ± 12.4 eu), and the entropy increments for the 10 nearest-neighbors are more favorable by 3.5 eu, on average.

Error Analysis. To evaluate the relative uncertainties in the nearest-neighbor parameters derived above, we determined how the experimental errors, σ_i, propagated to the errors in the nearest neighbors, σ_j, as described [Press et al. (1989) eq 14.3.19]. In the regression analysis described above, the experimental data were weighted assuming 5%, 10%, and 10% uncertainties in the ΔG°_{37}, ΔH°, and ΔS°, respectively. The experimental uncertainties given in Table 1 were not used because they reflect the experimental reproducibility of the data (i.e., precision) not the accuracy of the data (Bevington, 1969). Data from the literature were assigned the same percent errors. Assigning the same percent error for all the data effectively weights the data for all

Table 1: Thermodynamic Parameters of Duplex Formation[a]

DNA duplex	$1/T_M$ vs log C_T parameters				curve fit parameters			
	$-\Delta G°_{37}$ (kcal/mol)	$-\Delta H°$ (kcal/mol)	$-\Delta S°$ (eu)	T_M (°C)[b]	$-\Delta G°_{37}$ (kcal/mol)	$-\Delta H°$ (kcal/mol)	$-\Delta S°$ (eu)	$-\Delta C_p$ [kcal/(K mol)][c]
Two-State Transitions								
CCGG	3.4 ± 0.3	31.9 ± 1.1	91.9 ± 2.5	16.7	3.6 ± 0.1	29.4 ± 1.0	83.0 ± 3.5	0.2
CGCG	3.9 ± 0.4	38.5 ± 1.6	111.5 ± 3.8	23.6	4.2 ± 0.2	34.0 ± 1.7	96.2 ± 6.0	0.3
GCGC	4.3 ± 0.5	45.6 ± 2.4	133.3 ± 6.1	27.8	4.5 ± 0.1	35.8 ± 0.9	100.9 ± 2.4	
CGATCG	5.4 ± 0.4	33.1 ± 1.1	89.3 ± 2.3	34.3	5.3 ± 0.2	41.8 ± 3.8	117.9 ± 11.8	0.5
GACGTC	5.6 ± 1.3	37.2 ± 4.1	101.8 ± 9.2	36.4	5.5 ± 0.2	46.6 ± 3.8	132.6 ± 11.8	0.5
GCTAGC	5.4 ± 0.6	35.7 ± 1.8	97.7 ± 3.9	34.3	5.3 ± 0.1	42.7 ± 3.9	120.5 ± 12.2	
GGATCC	5.1 ± 0.2	30.2 ± 0.6	80.9 ± 1.3	30.8	4.9 ± 0.2	39.4 ± 3.8	111.4 ± 11.8	
CAAGCTTG	7.0 ± 0.2	54.7 ± 0.7	153.8 ± 1.8	44.6	7.0 ± 0.1	53.8 ± 0.8	150.9 ± 2.5	
CATCGATG	7.5 ± 0.6	56.3 ± 2.3	157.4 ± 5.6	47.4	7.7 ± 0.1	62.0 ± 2.4	175.1 ± 7.4	
CGATATCG	6.8 ± 0.2	51.8 ± 0.6	145.1 ± 1.4	44.0	6.9 ± 0.1	55.7 ± 3.9	157.1 ± 12.1	0.5
GAAGCTTC	6.9 ± 0.2	44.1 ± 0.8	120.3 ± 1.8	45.5	7.1 ± 0.2	54.7 ± 3.2	153.5 ± 9.8	0.5
GATCGATC	7.5 ± 0.6	53.6 ± 2.2	148.7 ± 5.2	47.7	7.7 ± 0.3	63.2 ± 5.9	178.8 ± 17.9	1.2
GATGCATC	7.2 ± 0.3	52.0 ± 1.1	144.2 ± 2.5	46.7	7.4 ± 0.2	60.3 ± 2.6	170.4 ± 7.9	
GGAATTCC	6.8 ± 0.2	46.1 ± 0.5	127.0 ± 1.2	44.5	6.8 ± 0.3	57.0 ± 2.7	161.8 ± 8.3	0.4
GGACGTCC	8.9 ± 0.2	58.6 ± 0.8	160.2 ± 1.8	55.2	9.1 ± 0.1	61.4 ± 0.8	168.7 ± 2.4	
GGAGCTCC	8.6 ± 0.3	53.4 ± 0.9	144.6 ± 2.1	54.7	8.8 ± 0.2	60.1 ± 1.8	165.4 ± 5.1	
GTACGTAC	7.0 ± 0.2	51.3 ± 0.6	142.7 ± 1.5	45.1	7.1 ± 0.2	56.4 ± 3.7	159.0 ± 11.4	0.7
GTAGCTAC	7.0 ± 0.1	51.4 ± 0.6	143.3 ± 1.3	44.9	7.0 ± 0.1	55.1 ± 2.3	155.0 ± 7.0	0.5
GTTGCAAC	7.3 ± 0.5	47.8 ± 1.6	130.5 ± 3.8	48.2	7.6 ± 0.2	59.7 ± 1.2	168.1 ± 3.7	
CCATCGCTACC/GGTAGCGATGG	13.5 ± 0.3	86.9 ± 1.2	236.8 ± 3.0	63.9	13.2 ± 0.2	82.8 ± 1.8	224.5 ± 5.2	
CCATTGCTACC/GGTAACGATGG	12.3 ± 0.3	83.9 ± 1.4	231.1 ± 3.4	59.7	12.1 ± 0.1	81.1 ± 0.6	222.4 ± 1.7	
CTGACAAGTGTC/GACTGTTCACAG	13.0 ± 0.6	88.6 ± 2.8	243.7 ± 6.9	61.6	12.1 ± 0.1	74.3 ± 2.2	200.6 ± 6.8	
CATATGGCCATATG	13.4 ± 0.5	93.7 ± 2.2	258.7 ± 5.6	65.0	12.0 ± 0.3	75.4 ± 2.4	204.3 ± 7.0	
Non-Two-State Transitions								
GTATACCGGTATAC	13.3 ± 0.1	101.7 ± 0.7	285.3 ± 1.9	61.9	11.3 ± 0.2	74.3 ± 0.7	203.0 ± 1.9	
CATATTGGCCAATATG	14.9 ± 1.1	110.2 ± 5.8	307.4 ± 15.1	65.3	11.6 ± 0.3	67.3 ± 3.0	179.7 ± 8.9	
GTATAACCGGTTATAC	15.3 ± 0.3	113.6 ± 1.7	316.8 ± 4.3	65.9	13.2 ± 0.3	85.8 ± 1.6	234.0 ± 4.5	

[a] Listed by oligomer length and in alphabetical order. For self-complementary sequences, only the top strand is given. For non-self-complementary duplexes, both strands are given in antiparallel orientation separated by a slash. Solutions are 1 M NaCl, 10 mM sodium cacodylate, and 0.5 mM Na_2EDTA, pH 7. Errors are standard deviations from the regression analysis of the melting data. Extra significant figures are given for $\Delta H°$ and $\Delta S°$ to allow accurate calculation of $\Delta G°_{37}$ and T_M. [b] Calculated for 10^{-4} M oligomer concentration for self-complementary sequences and 4×10^{-4} M for non-self-complementary sequences. [c] Only those sequences with a $\Delta H°$ vs T_M plot with a correlation coefficient greater than 0.8 are listed. Errors in ΔC_p are approximately 50%.

sequences equally in the regression analysis. Thus, the percent error assumed has no effect on the values of the nearest-neighbor parameters obtained, only on the propagated errors in the parameters. For example, if we assume errors of 5% for $\Delta H°$, we obtain the same nearest-neighbor parameters as with 10% errors, but the error estimates for the nearest neighbors are half as large. The nearest-neighbor errors, σ_i, given in Table 3 are the standard deviations that resulted from the propagation of experimental errors, σ_i, during linear regression. The free-energy covariances (Press et al., 1989) between pairs of nearest neighbors are small (less than ±0.002 kcal^2/mol^2 covariance) and can be neglected. The initiation parameter, however, covaries with all 10 of the nearest neighbors (−0.006 kcal^2/mol^2 covariance, on average).

Goodness of the fits for the nearest neighbor parameters for $\Delta G°_{37}$, $\Delta H°$, and $\Delta S°$ were evaluated from the values of χ^2 (eq 6) and the probability Q (Press et al., 1989). Q is the probability that a χ^2 larger than that observed could be obtained by chance (larger Q indicates a better fit). Q probabilities greater than 0.001 are considered statistically acceptable (Press et al., 1989). Q was calculated from the incomplete gamma function, gamma [ν/2, χ^2/2], where ν is the number of degrees at freedom (ν = number of independent measurements minus the number of parameters derived from the data) (Press et al., 1989). Since measurements were made on 44 different oligonucleotides and 11 parameters (10 nearest neighbors plus initiation) were determined, ν is 33 (for $\Delta G°_{37}$) or 34 (for $\Delta H°$ and $\Delta S°$ which do not have the initiation parameter floating). For $\Delta G°_{37}$, $\chi^2 = 28.4$ and $Q = 0.70$. For $\Delta H°$, $\chi^2 = 33.8$ and $Q = 0.47$. For $\Delta S°$, $\chi^2 = 50.7$ and $Q = 0.03$. These results suggest that within the estimated errors the nearest-neighbor model is a valid description of DNA thermodynamics in agreement with previous results (Sugimoto et al., 1994; Doktycz et al., 1995).

Comparison of Experimental vs Predicted Thermodynamics. Table 2 compares the experimental results for 60 oligonucleotides with those predicted using the nearest-neighbor parameters in Table 3. The 44 sequences that have two-state transitions are well predicted by the parameters in Table 3. For the T_M at 0.1 mM, the largest difference is 4.6 °C with an average deviation of 1.8 °C. The ability of the parameters in Table 3 to predict the T_M accurately is encouraging, since the parameters were not specifically optimized for the prediction of the T_M. The average deviations between experiment and prediction for $\Delta G°_{37}$, $\Delta H°$, and $\Delta S°$ are 4%, 7%, and 8%, respectively. Previous results for RNA (Kierzek et al., 1986), DNA (Sugimoto et al., 1994), and DNA/RNA hybrid (Sugimoto et al., 1995) oligonucleotides with different sequences, but the same nearest neighbors, suggest the nearest-neighbor model should be able to predict $\Delta G°_{37}$, $\Delta H°$, and T_M (at 0.1 mM) with average deviations of roughly 6%, 8%, and 2 °C, respectively. Thus, the predictive capacity of the parameters in Table 3 is within the limits of the nearest-neighbor model. The χ^2 and Q parameters, described above, also suggest that the nearest-neighbor parameters given in Table 3 adequately

Table 2: Experimental and Predicted Thermodynamic Parameters of Duplex Formation[a]

		experimental				predicted			
sequence	ref[b]	$-\Delta G°_{37}$ (kcal/mol)	$-\Delta H°$ (kcal/mol)	$-\Delta S°$ (eu)	T_M (°C)[c]	$-\Delta G°_{37}$ (kcal/mol)	$-\Delta H°$ (kcal/mol)	$-\Delta S°$ (eu)	T_M (°C)[c]
Molecules with Two-State Thermodynamics									
CCGG		3.5	30.6	87.4	16.6	3.4	23.5	64.0	12.4
CGCG		4.0	36.3	103.9	23.7	4.2	31.3	86.7	25.0
GCGC		4.4	40.7	117.1	27.5	4.4	32.3	89.6	26.2
CCGCGG	*d*	8.0	41.4	107.8	55.2	7.8	44.7	117.9	55.0
CGATCG		5.3	37.5	103.6	34.3	5.6	42.1	117.7	36.4
CGCGCG	*a*	8.3	46.4	122.8	55.7	8.6	52.5	140.6	57.3
CGGCCG	*d*	8.3	38.7	98.0	59.6	7.8	44.7	117.9	55.0
CGTACG	*f*	5.4	45.7	130.0	35.0	5.4	43.7	122.8	36.6
GACGTC		5.6	41.9	117.2	36.1	5.7	42.7	119.4	36.9
GCATGC	*g*	5.6	42.2	118.0	36.5	5.8	45.5	121.5	38.0
GCCGGC	*d*	8.5	45.2	118.3	57.7	8.0	45.7	120.8	55.4
GCGAGC/CGCTCG	*h*	7.7	51.4	124.0	33.2	7.5	40.7	115.1	32.0
GCGCGC	*e*	9.1	39.6	162.7	56.1	8.8	53.5	143.5	57.5
GCTAGC		5.3	39.2	109.1	34.3	5.3	40.7	114.8	32.6
GGATCC		5.0	34.8	96.2	30.8	5.0	35.3	97.9	30.6
GGCGCC	*d*	7.9	43.5	114.7	53.9	8.0	45.7	120.8	55.4
GGGACC/CCCTGG	*h*	6.5	32.7	84.5	44.9	6.4	36.4	96.0	45.3
GTGAAC/CACTTG	*h*	5.1	43.6	124.0	33.2	4.9	40.7	115.1	32.0
CAAAAAG/GTTTTTC	*i*	4.8	47.0	136.0	31.5	4.8	47.1	135.7	32.7
CAAAAAAG/GTTTTTTC	*j*	5.7	59.0	172.0	36.9	5.8	55.5	159.3	39.4
CAAGCTTG		7.0	54.2	152.3	44.6	7.2	54.9	153.7	46.0
CATCGATG		7.6	59.2	166.3	47.4	7.0	53.3	149.6	44.3
CGATATCG		6.9	53.7	151.1	44.1	6.9	54.9	155.0	43.6
CGTCGACG	*k*	9.8	64.1	175.1	58.3	9.8	62.9	170.4	60.2
GAAGCTTC		7.0	49.4	136.9	45.2	7.3	55.5	155.7	45.8
GATCGATC		7.6	58.4	163.7	47.6	7.2	53.9	151.6	44.1
GATGCATC		7.3	56.1	157.3	46.5	7.2	54.3	152.5	44.8
GGAATTCC		6.8	51.6	144.4	43.8	7.0	52.1	145.1	45.7
GGACGTCC		9.0	60.0	164.5	55.1	9.2	56.1	150.6	59.0
GGAGCTCC		8.7	56.8	155.0	54.4	8.8	52.1	139.7	56.6
GGTATACC	*k*	5.5	54.5	158.0	36.0	6.1	49.7	140.3	40.2
GTACGTAC		7.1	53.8	150.9	45.1	6.8	57.1	161.8	43.9
GTAGCTAC		7.0	53.3	149.2	44.9	6.4	53.1	150.9	40.7
GTTGCAAC		7.5	53.8	149.3	47.7	7.7	59.9	167.5	49.2
CAAAAAAAG/GTTTTTTTC	*i*	7.2	68.0	196.0	44.2	6.8	63.9	182.9	44.4
CAAACAAAG/GTTTGTTTC	*i*	7.7	64.5	183.0	47.3	7.6	63.1	178.0	48.3
CAAATAAAG/GTTTATTTC	*i*	6.5	58.6	168.0	41.4	6.1	59.0	173.0	40.9
CAAAGAAAG/GTTTCTTTC	*i*	7.3	62.8	179.0	45.2	7.4	60.9	172.1	46.7
GCGAATTCGC	*k*	12.9	80.0	216.3	67.9	12.2	81.8	221.7	64.8
CCATCGCTACC/GGTAGCGATGG		13.3	84.8	230.7	67.6	12.9	77.2	207.0	69.5
GCGAAAAGCG/CGCTTTTCGC	*l*	11.9	74.2	201.0	65.2	12.6	81.4	220.9	67.2
CCATTGCTACC/GGTAACGATGG		12.2	82.5	226.8	63.5	11.7	75.2	204.1	65.0
CTGACAAGTGTC/GACTGTTCACAG		12.6	81.5	222.1	65.7	12.9	84.0	229.2	66.2
CATATGGCCATATG		12.7	84.5	231.5	65.3	13.2	92.7	256.3	66.4
Molecules with Terminal A-T Base Pairs[n]									
TCATGA	*d*	3.3	30.4	152.0	22.8	3.4	35.9	105.3	17.3
TGATCA	*d*	2.8	52.6	160.6	20.9	3.4	35.9	105.3	17.3
AAAAAAAA/TTTTTTTT	*n*	4.5	59.8	178.2	31.2	3.9	58.4	171.1	35.2
TAGATCTA	*d*	5.1	49.2	142.2	33.4	4.2	45.9	135.9	24.5
TCTATAGA	*d*	4.3	45.7	133.4	28.1	4.2	45.9	135.9	24.5
ATGAGCTCAT	*o*	10.0	68.0	187.0	58.1	9.5	66.5	184.7	54.4
TTTTATAATAAA/AAAATATTATTT	*p*	5.5	75.6	226.0	36.3	5.8	81.5	240.6	41.6
CAACTTGATATTATTA/GTTGAACTATAATAAT	*q*	12.4	102.0	289.0	58.8	12.7	108.8	310.2	58.1
Molecules with Non-Two-State Thermodynamics									
CCCGGG	*r*	6.9	62.0	177.8	43.0	7.0	36.9	95.2	52.0
CCCAGGG/GGGTCCC	*r*	7.9	63.0	177.8	48.1	7.8	40.3	103.7	57.2
CGCGAATTCGCG	*s*	20.6	135.0	369.0	75.4	16.4	101.3	272.7	75.0
CGCATGGGTACGC/GCGTACCCATGCG	*t*	14.4	101.0	279.1	66.5	17.4	100.6	266.8	79.7
GTATACCGGTATAC		12.3	88.0	244.1	62.2	13.0	96.1	267.6	63.0
CATATTGGCCAATATG		13.2	88.8	243.6	65.9	15.3	109.5	303.5	67.1
GTATAACCGGTTATAC		14.3	99.7	275.4	66.3	15.0	112.9	314.8	65.8
CGCGTACGCGTACGCG	*u*	29.1	158.0	415.6	91.0	24.1	140.9	374.5	85.6

[a] Listed by oligomer length and in alphabetical order. For self-complementary sequences only the top strand is given. For non-self-complementary sequences both strands are given in antiparallel orientation separated by a slash. [b] Sequences without a literature reference are from Table 1 of this work. [c] Calculated for 10^{-4} M oligomer concentration for self-complementary sequences and 4×10^{-4} M for non-self-complementary sequences. [d] Sugimoto et al. (1994). [e] Senior et al. (1988). [f] Breslauer (1986). [g] Williams et al. (1989). [h] Li and Agrawal (1995). [i] Aboul-ela et al. (1985). [j] Morden et al. (1983). [k] Breslauer et al. (1986). [l] LeBlanc and Morden (1991). [m] For each 5′-terminal T-A base pair, +0.4 kcal/mol is added to both $\Delta H°$ and $\Delta G°_{37}$. [n] Sugimoto et al. (1991). [o] Bolewska et al. (1984). [p] Tibanyenda et al. (1984). [q] Arnold et al. (1987). [r] Murky et al. (1983). [s] Plum et al. (1992). [t] Raap et al. (1985).

Table 3: Thermodynamic Parameters for DNA Helix Initiation and Propagation in 1 M NaCl[a]

propagation sequence	ΔH° (kcal/mol)	ΔS° (eu)	ΔG°_{37} (kcal/mol)
AATT	−8.4 ± 0.7	−23.6 ± 1.8	−1.02 ± 0.04
ATTA	−6.5 ± 0.8	−18.8 ± 2.3	−0.73 ± 0.05
TA/AT	−6.3 ± 1.0	−18.5 ± 2.6	−0.60 ± 0.05
CA/GT	−7.4 ± 1.1	−19.3 ± 2.9	−1.38 ± 0.06
GT/CA	−8.6 ± 0.7	−23.0 ± 2.0	−1.43 ± 0.05
CT/GA	−6.1 ± 1.2	−16.1 ± 3.3	−1.16 ± 0.07
GA/CT	−7.7 ± 0.7	−20.3 ± 1.9	−1.46 ± 0.05
CG/GC	−10.1 ± 0.9	−25.5 ± 2.3	−2.09 ± 0.07
GC/CG	−11.1 ± 1.0	−28.4 ± 2.6	−2.28 ± 0.08
GG/CC	−6.7 ± 0.6	−15.6 ± 1.5	−1.77 ± 0.06
initiation at G·C[b]	(0)	(−5.9 ± 0.8)	+1.82 ± 0.24
initiation at A·T[c]	(0)	(−9.0 ± 3.2)	(+2.8 ± 1)
symmetry correction[d]	0	−1.4	+0.4
5′-terminal T·A bp[e]	+0.4	0	+0.4

[a] Errors are standard deviations. Extra significant figures are given for ΔH° and ΔS° to allow accurate calculation of the T_M. Values in parentheses involve assumptions about the initiation process (see text). [b] Initiation parameter for duplexes that contain at least one G·C base pair. [c] Initiation parameter for duplexes that contain only A·T base pairs. [d] Symmetry correction applies only to self-complementary sequences. [e] To account for end effects, duplexes are given the penalty listed for each terminal 5′-T·A-3′ base pair. Note this penalty is not applied to sequences with terminal 5′-A·T-3′ base pairs (see text).

describe DNA thermal denaturation.

Applicability to Non-Two-State Transitions. We believe the parameters in Table 3 apply to duplexes from 4 to 20 base pairs. Beyond 20 base pairs, DNA transitions are unlikely to be two-state. Transitions that are not two-state require a statistical mechanical model for accurate predictions (Gralla & Crothers, 1973; Steger, 1994). Table 2 lists eight oligonucleotides that are not two-state. When the ΔG°_{37}, ΔH°, and T_M of these oligomers are predicted with the parameters in Table 3, the average deviations of measured versus predicted values are 14%, 21%, and 4.7 °C, respectively. This suggests that the two-state model can also provide reasonable approximations for oligomers that do not have strictly two-state transitions.

DISCUSSION

Application of the Nearest-Neighbor Parameters. The nearest-neighbor model asserts that the free-energy for duplex formation is the sum of three terms: (1) an unfavorable entropy associated with the loss of translational freedom upon formation of the first hydrogen bonded base pair (i.e., the initiation free energy), (2) the sum of terms for the pairwise interactions between base pairs, and (3) an entropic penalty (Bailey & Monahan, 1978; Cantor & Schimmel, 1980) for the maintenance of the *C*2 symmetry of self-complementary duplexes (eq 4). Scheme 1 illustrates the calculation of ΔG°_{37} for the sequence GCTAGC using the parameters in Table 3. Similarly, the predicted enthalpy change for GCTAGC is: ΔH°(predicted) = 2(−11.1) + 2(−6.1) + (−6.3) = −40.7 kcal/mol. The measured value is −39.2 kcal/mol. Note that ΔH° for initiation and symmetry are zero. The predicted entropy change for GCTAGC is ΔS°(predicted) = 2(−28.4) + 2(−16.1) + (−18.5) − 5.9 − 1.4 = −114.8 eu. The measured value is −109.1 eu. The T_M is predicted at a given oligonucleotide concentration (C_T/4 for non-self-complementary sequences) using eq 2 along with the predicted ΔH° and ΔS°. For example, the predicted T_M for GCTAGC at 0.1 mM is T_M (predicted) = (−40 700 cal/mol)/(−114.8 eu + 1.987 eu × ln (1 × 10^{-4})) = 305.8 K = 32.6 °C. The measured value is 34.3 °C. Note that the units for ΔH° are kcal/mol and must be multiplied by 1000 to be consistent with ΔS° and the gas constant, R, which are in eu [cal/(K mol)].

Scheme 1. Prediction of ΔG°_{37}

```
          ↓  ↓  ↓
   5′ G-C-T-A-G-C 3′
      * * * * * *
   3′ C-G-A-T-C-G 5′
          ↑  ↑
```

ΔG°_{37}(predicted) = 2 ΔG°_{37}(GC/CG) + 2 ΔG°_{37}(CT/GA) + ΔG°_{37}(TA/AT)
+ ΔG°_{37}(init) + ΔG°_{37}(sym)
= 2 (-2.28) + 2(-1.16) + (-0.60) + 1.82 + 0.4
ΔG°_{37}(predicted) = -5.26 kcal/mol.
ΔG°_{37}(observed) = -5.3 kcal/mol.

Trends in the Nearest-Neighbor Parameters. The observed trend in nearest-neighbor stabilities at 37 °C is GC > CG > GG > GA ≈ GT ≈ CA > CT > AA > AT > TA (where only the top strand is shown for each nearest neighbor). This trend suggests that both sequence and base composition are important determinants of DNA duplex stability. It has long been recognized that DNA stability depends of the percent G-C content (Marmur & Doty, 1962). The ΔG°_{37} parameters in Table 3 show that there are significant sequence dependent contributions superimposed on the general trend. On the other hand, the nearest-neighbor ΔH° parameters (Table 3) do not follow this trend. This suggests that stacking, hydrogen bonding, and other contributions to the ΔH° have a complicated sequence dependence. Perhaps, this is not surprising since it is well known that the detailed structure of DNA is profoundly dependent on sequence (Callidine & Drew, 1984; Hunter, 1993).

The average of the ΔS°'s for the 10 nearest neighbor propagations is −20.9 eu. The agrees reasonably well with the sequence independent value of −24.85 ± 1.74 eu/base pair derived from polymers dissolved in 0.075 M Na^+ (Delcourt & Blake, 1991). Our results are also consistent with a simplistic calculation of the conformational entropy (Cantor & Schimmel, 1980):

$$\Delta S^\circ_{conf} = 2R \ln(3 \times 7 \times 2 \times 3 \times 3 \times 2) = -26.3 \text{ eu/base pair}$$

where the numbers inside the parentheses are the assumed number of possible conformations for the α, $\beta\gamma$ (together), δ, ε, ζ, and χ dihedral angles. The "2" in front of the gas constant is required because two residues must be constrained to form a base pair and propagate a helix. This calculation systematically overestimates ΔS°_{conf} because many of the possible conformations would have high energies associated with them (due to steric repulsion). A more rigorous calculation would weight each of the possible conformations with a Boltzmann factor. This calculation also neglects salt effects and hydrophobic contributions to stacking (Hunter, 1993).

Comparison with Previous DNA Nearest-Neighbor Parameters. Breslauer et al. (1986) derived nearest-neighbor parameters from a data set of 19 oligonucleotides (dissolved in 1 M NaCl) and nine polymers (dissolved in low salt with

results extrapolated to high salt) and assumed a value of 5.2 kcal/mol for helix initiation (Borer et al., 1974). Our results share some similarities but also differ significantly from those of Breslauer and co-workers. Except for the GG/CC nearest neighbor, the ΔH° values of Breslauer et al. are within 2 kcal/mol of those in Table 3 (for GG/CC Breslauer reports $\Delta H^\circ = -11.0$ kcal/mol, whereas we observe -6.7 kcal/mol). The trend for the stabilities of nearest neighbors with only AT base pairs is also similar with AA/TT > AT/TA > TA/AT. On the other hand, the parameters of Breslauer et al. (1986) have the stability order CG/GC > GC/CG ≈ GG/CC, while GC/CG > CG/GC > GG/CC in our parameter set. Breslauer et al. also have CA/GT > AA/TT > GA/CT ≈ CT/GA > GT/CA, while GA/CT ≈ GT/CA ≈ CA/GT > CT/GA > AA/TT in our parameter set. On average, Breslauer's parameters predict the ΔG°_{37}, ΔH°, ΔS°, and T_M of the two-state molecules given in Table 2 with average deviations of 16%, 12%, 13%, and 6.0 °C, respectively. As discussed above, the parameters in Table 3 predict the ΔG°_{37}, ΔH°, ΔS°, and T_M of the two-state molecules given in Table 2 with average deviations of 4%, 7%, 8%, and 1.8 °C, respectively. This comparison is somewhat biased since our parameters were optimized to predict our database, while Breslauer's parameters were derived from an independent data set. Quartin and Wetmur (1989) use essentially the same data as Breslauer et al. (1986) but assume a value of +2.2 kcal/mol for helix initiation (Pohl, 1974). These parameters predict the ΔG°_{37} and the T_M of the two-state molecules in Table 2 with average deviations of 10% and 4.5 °C, respectively.

Comparison with RNA Nearest-Neighbor Parameters. Our DNA parameters also differ significantly from RNA parameters measured by Freier et al. (1986a). This is not surprising, however, because RNA and DNA helices are known to have different structures (i.e., A-form vs B-form). Some DNA nearest neighbors are more stable while others are less stable than the analogous RNA nearest neighbors. For example, the DNA nearest neighbors AA/TT and CG/GC are slightly more stable than the corresponding RNA nearest neighbors AA/UU and CG/GC. In all the other cases, DNA nearest neighbors are less stable than RNA. We also observe that DNA nearest neighbors with only C-G base pairs are less sequence dependent (largest difference = 0.51 kcal/mol) than the corresponding RNA nearest neighbors (largest difference 1.4 kcal/mol). The relative order for stabilities of nearest neighbors with only A-T (DNA) or A-U (RNA) base pairs are also different with AA/TT > AT/TA > TA/AT (DNA) vs UA/AU > AU/UA = AA/UU (RNA). Apparently, the relative stability of DNA and RNA duplexes depends on base sequence.

The helix initiation parameter is more favorable in DNA (+1.82 kcal/mol) than in RNA (+3.4 kcal/mol). This is somewhat puzzling as DNA and RNA duplex initiation are expected to be similar since both require two strands to associate and reduce their translational and rotational degrees of freedom by forming a hydrogen-bonded base pair. Further work is required to elucidate the origin of this effect.

Salt Dependence. The parameters in Table 3 apply to oligomers dissolved in 1 M NaCl at pH 7. To allow for approximate predictions in lower salt environments, we suggest the following preliminary equation:

$$T_M^{[Na]} = T_M^{1M\ Na} + 12.5 \log [Na^+] \qquad (7)$$

where $T_M^{1M\ Na}$ is T_M predicted from Table 3 (1 M NaCl), and $T_M^{[Na]}$ is the T_M predicted at the desired sodium concentration. This correction for the T_M is in agreement with that determined previously (Erie et al., 1987; Rentzeperis et al., 1993) for oligonucleotides but is somewhat smaller than that observed in polymers (Marmur & Doty, 1962; Schildkraut & Lifson, 1965). Between 0.1 and 1 M NaCl, this correction predicts the T_M of 26 sequences (dissolved in 0.1–0.3 M NaCl) from the literature (Aboul-ela et al., 1985; Braunlin & Bloomfield, 1991; Gaffney & Jones, 1989; Kawase et al., 1986; Williams et al., 1989; Lesnik & Freier, 1995) with an average deviation 3.5 °C (Allawi and SantaLucia, unpublished results). Below 0.1 M, this correction is not reliable. This correction assumes that trends in nearest-neighbor stability are independent of salt concentration. Counterion-condensation theory suggests this assumption is reasonable since the salt behavior depends on the spacing between phosphates which should be relatively independent of sequence (Manning, 1978). However, this theory applies strictly to polymers and salt concentration below 0.1 M, and, for short oligonucleotides, the salt behavior may depend on oligonucleotide length (Record & Lohman, 1978; Olmsted et al., 1989). Two experimental studies, however, suggest that the salt behavior of oligomers is remarkably similar to that of polymers (Williams et al., 1989; Braunlin & Bloomfield, 1991). While the above corrections were derived for sodium counterions, potassium counterions probably follow the same trend. The behavior of oligonucleotide thermodynamics in the presence of divalent cations, however, is likely to be more complicated. Previous work indicates that 1 M NaCl mimics 0.15 M NaCl/10 mM $MgCl_2$ (Williams et al., 1989)—a condition similar to those commonly used in PCR reactions. Clearly, further work on the salt dependence of oligonucleotide thermal denaturation is required (Kumar, 1995).

ACKNOWLEDGMENT

We thank David Hyndman (Advanced Gene Computing Technologies) for stimulating conversations and Mieko Ogura (Hitachi Chemical Research) for synthesizing oligonucleotides. We thank Jeff McDowell and Douglas H. Turner for providing the program MELTWIN v2.1 for the analysis of optical melting curves. We thank Martin McClain for instruction in using MATHEMATICA (Wolfram Research) for linear regression analysis. We also thank Douglas H. Turner, Philip N. Borer, and Kenneth J. Breslauer for critical reading of the manuscript.

SUPPORTING INFORMATION AVAILABLE

Six figures showing $1/T_M$ vs $\ln C_T$ plots for the 20 sequences presented in Table 1 which are not shown in Figure 1 (3 pages). Ordering information is given on any current masthead page.

REFERENCES

Aboul-ela, F., Koh, D., Tinoco, I., Jr., & Martin, F. H. (1985) *Nucleic Acids Res. 13*, 4811–4824.

Aida, M. (1988) *J. Theor. Biol. 130*, 327–335.

Albergo, D. D., Marky, L. A., Breslauer, K. J., & Turner, D. H. (1981) *Biochemistry 20*, 1409−1413.
Arnold, F. H., Wolk, S., Cruz, P., & Tinoco, I., Jr. (1987) *Biochemistry 26*, 4068−4075.
Bailey, W. F., & Monahan, A. S. (1978) *J. Chem. Ed. 55*, 489−493.
Bevington, P. R. (1969) *Data Reduction and Error Analysis for the Physical Sciences*, pp 164−186 and 187−203, McGraw-Hill, New York.
Bolewska, K., Zielenkiewicz, A., & Wierzchowski, K. L. (1984) *Nucleic Acids Res. 12*, 3245−3256.
Borer, P. N., Dengler, B., Tinoco, I., Jr., & Uhlenbeck, O. C. (1974) *J. Mol. Biol. 86*, 843−853.
Braunlin, W. H., & Bloomfield, V. A. (1991) *Biochemistry 30*, 754−758.
Breslauer, K. J. (1986) in *Thermodynamic Data for Biochemistry and Biotechnology* (Hinz, H., Ed.) pp 402−427, Springer-Verlag, New York.
Breslauer, K. J., Frank, R., Blocker, H., & Marky, L. A. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 3746−3750.
Brown, T., & Brown, D. J. S. (1991) in *Oligonucleotides and Analogues* (Eckstein, F., Ed.) pp 1−24, IRL Press, New York.
Callidine, C. R., & Drew, H. R. (1984) *J. Mol. Biol. 178*, 773−782.
Cantor, C. R., & Schimmel, P. R. (1980) *Biophysical Chemistry Part III: The Behavior of Biological Macromolecules*, pp 1183−1264, W. H. Freeman, San Francisco, CA.
Chou, S.-H., Flynn, P., & Reid, B. (1989) *Biochemistry 28*, 2422−2435.
Delcourt, S. G., & Blake, R. D. (1991) *J. Biol. Chem. 266*, 15160−15169.
Doktycz, M. J., Goldstein, R. F., Paner, T. M., Gallo, F. J., & Benight, A. S. (1992) *Biopolymers 32*, 849−864.
Doktycz, M. J., Morris, M. D., Dormady, S. J., Beattie, K. L., & Jacobson, K. B. (1995) *J. Biol. Chem. 270*, 8439−8445.
Erie, D., Sinha, N., Olson, W., Jones, R., & Breslauer, K. (1987) *Biochemistry 26*, 7150−7159.
Fodor, S. P. A., Rava, R. P., Huang, X. C., Pease, A. C., Holmes, C. P., & Adams, C. L. (1993) *Nature 364*, 555−556.
Freier, S. M., Kierzek, R., Jaeger, J. A., Sugimoto, N., Caruthers, M. H., Neilson, T., & Turner, D. H. (1986a) *Proc. Natl. Acad. Sci. U.S.A. 83*, 9373−9377.
Freier, S. M., Sugimoto, N., Sinclair, A., Alkema, D., Neilson, T., Kierzek, R., Caruthers, M. H., & Turner, D. H. (1986b) *Biochemistry 25*, 3214−3219.
Freier, S. M. (1993) in *Antisense Research and Applications* (Crooke, S. T., & Lebleu, B., Eds.) pp 67−82, CRC Press, Boca Raton, FL.
Gaffney, B. L., & Jones, R. A. (1989) *Biochemistry 28*, 5881−5889.
Gotoh, O., & Tagashira, Y. (1981) *Biopolymers 20*, 1033−1042.
Gralla, J., & Crothers, D. M. (1973) *J. Mol. Biol. 78*, 301−319.
Gray, D. M., & Tinoco, I., Jr. (1970) *Biopolymers 9*, 223−244.
Hunter, C. A. (1993) *J. Mol. Biol. 230*, 1025−1054.
Kawase, Y., Iwai, S., Inoue, H., Miura, K., & Ohtsuka, E. (1986) *Nucleic Acids Res. 14*, 7727−7736.
Kierzek, R., Caruthers, M. H., Longfellow, C. E., Swinton, D., Turner, D. H., & Freier, S. M. (1986) *Biochemistry 25*, 7840−7846.
Klump, H. H. (1990) in Landolt-Bornstein, New series, VII *Biophysics*, Vol. I, *Nucleic Acids*, Subvol. c, *Spectroscopic and Kinetic Data*, Physical Data I. (Saenger, W., Ed.) pp 241−256, Springer-Verlag, Berlin.
Kumar, A. (1995) *Biochemistry 34*, 12921−12925.
LeBlanc, D. A., & Morden, K. M. (1991) *Biochemistry 30*, 4042−4047.
Lesnik, E. A., & Freier, S. M. (1995) *Biochemistry 34*, 10807−10815.
Li, Y., & Agrawal, S. (1995) *Biochemistry 34*, 10056−10062.
Li, Y., Zon, G., & Wilson, W. D. (1991) *Biochemistry 30*, 7566−7572.
Manning, G. (1978) *Q. Rev. Biophys. 11*, 179−246.
Marky, L. A., & Breslauer, K. J. (1987) *Biopolymers 26*, 1601−1620.
Marky, L. A., Blumenfeld, K. S., Kozlowski, S., & Breslauer, K. J. (1983) *Biopolymers 22*, 1247−1257.
Mendelman, L. V., Boosalis, M. S., Petruska, J., & Goodman, M. F. (1989) *J. Biol. Chem. 264*, 14415−14423.
Marmur, J., & Doty, P. (1962) *J. Mol. Biol. 5*, 109−118.
Morden, K. M., Chu, Y. G., Martin, F. H., & Tinoco, I., Jr. (1983) *Biochemistry 21*, 428−436.
Olmsted, M. C., Anderson, C. F., & Record, M. T., Jr. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 7766−7770.
Ornstein, R., & Fresco, J. R. (1983) *Biopolymers 22*, 1979−2000.
Otto, P. (1989) *J. Mol. Struct. 188*, 277−288.
Petersheim, M., & Turner, D. H. (1983) *Biochemistry 22*, 256−263.
Petruska, J., Goodman, M. F., Boosalis, M. S., Sowers, L. S., Cheong, C., & Tinoco, I., Jr. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 6252−6256.
Plum, G. E., Grollman, A. P., Johnson, F., & Breslauer, K. J. (1992) *Biochemistry 31*, 12096−12102.
Pohl, F. M. (1974) *Eur. J. Biochem. 42*, 495−504.
Press, W. H., Flannery, B. P., Teukolsky, S. A., & Vetterling, W. T. (1989) *Numerical Recipes*, pp 52−64 and 498−520, Cambridge University Press, New York.
Quartin, R. S., & Wetmur, J. G. (1989) *Biochemistry 28*, 1040−1047.
Raap, J., van der Marel, G. A., van Boom, J. H., Joordens, J. J. M., & Hilbers, C. W. (1985) in *Fourth Conversation in Biomolecular Stereodynamics* (Sarma, R. H., Ed.) p 122a, Adenine Press, Guilderland, NY.
Record, M. T., Jr., & Lohman, T. M. (1978) *Biopolymers 17*, 159−166.
Renzeperis, D., Ho, J., & Marky, L. A. (1993) *Biochemistry 32*, 2564−2572.
Richards, E. G. (1975) in *Handbook of Biochemistry and Molecular Biology: Nucleic Acids* (Fasman, G. D., Ed.) 3rd ed., Vol. 1, p 597, CRC Press, Cleveland, OH.
Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S., Higuchi, R. H., Horn, G. T., Mullis, K. B., & Erlich, H. A. (1988) *Science 239*, 487−494.
SantaLucia, J., Jr., Kierzek, R., & Turner, D. H. (1990) *Biochemistry 29*, 8813−8819.
Scheffler, I. E., Elson, E. L., & Baldwin, R. L. (1970) *J. Mol. Biol. 48*, 145−171.
Schildkraut, C., & Lifson, S. (1965) *Biopolymers 3*, 195−208.
Senior, M., Jones, R. A., & Breslauer, K. J. (1988) *Biochemistry 27*, 3879−3885.
Southern, E. M. (1975) *J. Mol. Biol. 98*, 503−517.
Steger, G. (1994) *Nucleic Acids Res. 22*, 2760−2768.
Sugimoto, N., Tanaka, A., Shintani, Y., & Sasaki, M. (1991) *Chem. Lett.* 9−12.
Sugimoto, N., Honda, K., & Sasaki, M. (1994) *Nucleosides Nucleotides 13*, 1311−1317.
Sugimoto, N., Nakano, S., Katoh, M., Matsumura, A., Nakamuta, H., Ohmichi, T., Yonegama, M., & Sasaki, M. (1995) *Biochemistry 34*, 1121−11216.
Tibanyenda, N., De Bruin, S. H., Haasnoot, C. A. G., van der Marel, G. A., van Boom, J. H., & Hilbers, C. W. (1984) *Eur. J. Biochem. 139*, 19−27.
Turner, D. H., Sugimoto, N., & Freier, S. M. (1990) in Landolt-Bornstein, New series, VII *Biophysics*, Vol. I, *Nucleic Acids*, Subvol. e, *Spectroscopic and Kinetic Data*, Physical Data I. (Saenger, W., Ed.) pp 201−227, Springer-Verlag, Berlin.
Vologodskii, A. V., Amirikyan, B. R., Lyubchenko, Y. L., & Frank-Kamenetskii, M. D. (1984) *J. Biomol. Struct. Dyn. 2*, 131−148.
Wartell, R. M., & Benight, A. S. (1985) *Phys. Rep. 126*, 67−107.
Williams, A. P., Longfellow, C. E., Freier, S. M., Kierzek, R., & Turner, D. H. (1989) *Biochemistry 28*, 4283−4291.
Wolfram, S. (1992) *MATHEMATICA* version 2.1, Wolfram Research, Inc.

BI951907Q

Published online 3 May 2007 *Nucleic Acids Research, 2007, Vol. 35, No. 10 3367–3374*
doi:10.1093/nar/gkm210

On the stability of peptide nucleic acid duplexes in the presence of organic solvents

Anjana Sen and Peter E. Nielsen*

Department of Cellular and Molecular Medicine, Faculty of Health Sciences, University of Copenhagen, The Panum Institute, Blegdamsvej 3c, DK-2200 Copenhagen N, Denmark

Received January 11, 2007; Accepted March 27, 2007

ABSTRACT

Nucleic acid double helices are stabilized by hydrogen bonding and stacking forces (a combination of hydrophobic, dispersive and electrostatic forces) of the base pairs in the helix. One would predict the hydrogen bonding contributions to increase and the stacking contributions to decrease as the water activity in the medium decreases. Study of nucleobase paired duplexes in the absence of water and ultimately in pure aprotic, non-polar organic solvents is not possible with natural phosphodiester nucleic acids due to the ionic phosphate groups and the associated cations, but could be possible with non-ionic nucleic acid analogues or mimics such as peptide nucleic acids. We now report that peptide nucleic acid (PNA) (in contrast to DNA) duplexes show almost unaffected stability in up to 70% dimethylformamide (DMF) or dioxane, and extrapolation of the data to conditions of 100% organic solvents indicates only minor (or no) destabilization of the PNA duplexes. Our data indicate that stacking forces contribute little if at all to the duplex stability under these conditions. The differences in behaviour between the PNA and the DNA duplexes are attributed to the differences in hydration and counter ion release rather than to the differences in nucleobase interaction. These results support the possibility of having stable nucleobase paired double helices in organic solvents.

INTRODUCTION

Nucleic acid double helices are stabilized by hydrogen bonding and stacking forces (a combination of hydrophobic, dispersive and electrostatic forces) of the base pairs in the helix (1–5). The most recent data suggest that stacking interactions are the more important for DNA duplex stability (6,7). Indeed these results suggest that while hydrogen bonding in GC base pairs are stabilizing the helix, hydrogen bonding in AT base pairs are slightly destabilizing (6,7). One would predict the hydrogen bonding contributions to increase and the stacking contributions to decrease as the water activity in the medium decreases. Therefore, it could be of great interest both from a structural and thermodynamic as well as from a functional point of view to study nucleobase paired duplexes in the absence of water and ultimately in pure aprotic, non-polar organic solvents.

This is not possible with natural phosphodiester nucleic acids due to the ionic phosphate groups and the associated cations, but could be possible with uncharged nucleic acid analogues or mimics such as methylphosphonates (8), phosphotriesters (9), morpholino derivatives (10) and peptide nucleic acids (PNAs) (11,12). PNAs are DNA pseudopeptide mimics capable of forming DNA-like Watson–Crick base-pair double helices with sequence complementary DNA, RNA and PNA oligomers (13–18). As the aminoethyl glycine backbone of PNA is charge neutral, no counter ions are required for stabilizing PNA-PNA duplexes, and consequently such duplexes may more conveniently be studied in organic solvents.

Results from a previous study has given preliminary indications that PNA–PNA duplexes are much less affected by the presence of organic co-solvent (50% DMF) than DNA–DNA duplexes, correlating well with the lack of measurable changes in hydration or counter ion binding upon PNA duplex formation (19). However, this observation warrants further exploration. We now report that PNA (in contrast to DNA) duplexes show almost unaffected stability in 70% dimethylformamide (DMF) or dioxane, and extrapolation of the data to conditions of 100% organic solvents indicates only a little (or no) destabilization of the PNA duplexes.

MATERIALS AND METHODS

PNAs

PNA1, PNA2, PNA3, PNA4 and PNA5 were synthesized using solid phase Boc chemistry, purified by HPLC and characterized by MALDI-TOF mass spectrometry as described previously (20). PNA concentrations

*To whom correspondence should be addressed. Tel: +45 35 327762; Fax: +45 35 396042; Email: pen@imbg.ku.dk

were determined spectrophotometrically at 65°C using molar extinction coefficients: ε_{260} of adenine = 15 400 $M^{-1}cm^{-1}$, ε_{260} of guanine = 11 700 $M^{-1}cm^{-1}$, ε_{260} of thymine = 8800 $M^{-1}cm^{-1}$ and ε_{260} of cytosine = 7400 $M^{-1}cm^{-1}$.

Chemicals and DNAs

All chemical reagents used were of analytical grade except for dimethylformamide (DMF) and dioxane, which were spectroscopic grade from Sigma-Aldrich, Munich, Germany. The DNA oligonucleotides were purchased from DNA Technology, Aarhus, Denmark, and used without further purification.

Sample preparation

Main stock solutions of PNAs and DNAs were prepared by dissolution in deionized, double distilled water. Experimental samples were made by diluting from the corresponding main stock solutions in 10 mM phosphate buffer (pH 7.2) containing 100 mM NaCl and 0.1 mM EDTA.

Equimolar mixtures (1:1 stoichiometry in single strands) of the PNA or DNA and its complementary strand were dissolved in the buffer mentioned above with desired amount of organic co-solvents. The duplex formation was assured by heating to 90°C and then cooling slowly to room temperature to allow proper annealing. No sign of aggregation or decreased solubility of the PNAs at up to 70% of organic co-solvents was observed.

UV-melting experiments

The thermal melting experiments were performed on a Cary 300 Bio UV-visible spectrophotometer (Varian, Cary, NC, USA) connected to a temperature controller. Thermal melting profiles were obtained using heating–cooling cycles in the range of −3 to 95°C. The melting temperature (T_m) was determined from the peak of the first derivative of the heating curve. Cuvettes of 1.0 cm path length and 1.0 ml volume were used for all experiments.

Thermal melting curves at >50% of DMF or dioxane start to lose the upper baseline and show severe disturbances partly because of high absorbance of DMF at the wavelength required for the experiments. Therefore, thermodynamic data at >50% of DMF could not be obtained. However, it was possible to obtain T_m values at 60 and 70% of DMF or dioxane.

Thermodynamics

The thermodynamic parameters viz enthalpy change (ΔH^0), entropy change (ΔS^0), and Gibbs' free energy change (ΔG^0) were evaluated using the 'hyperchromicity method' (curve fitting) and/or the concentration method (21).

The hyperchromicity method

The hyperchromicity method utilizes alpha curve and van't Hoff plots (ln K_T versus T^{-1}) according to the following definitions (21): The fraction (α_T) of single strands that remained hybridized in the duplex at a particular temperature T in Kelvin is represented as

$$\alpha_T = \frac{A_s - A}{A_s - A_d} \quad 1$$

where, A_d is the absorbance of the duplex in fully hybridized condition, A_s is the absorbance of the single strands in fully denatured condition and A_s is absorbance at a particular point on the thermal melting curve at temperature T.

For non-self-complementary sequences forming n-mer structures, the general equilibrium constant equation at a particular temperature T can be expressed as:

$$K_T = \frac{\alpha_T}{(1-\alpha_T)^n (c_{ts}/n)^{n-1}} \quad 2$$

where, c_{ts} represents the total concentration of strands and n is the molecularity of the complex. Assuming a two-state model, Equation (2) reduces to

$$K_T = \frac{2\alpha_T}{(1-\alpha_T)^2 c_{ts}} \quad 3$$

The van't Hoff plot ln K_T versus T^{-1} is a straight line represented by

$$\ln K_T = \left(-\frac{\Delta H^0}{R}\right)\frac{1}{T} + \left(\frac{\Delta S^0}{R}\right) \quad 4$$

Hence, ΔH^0 can be obtained from the slope and ΔS^0 can be obtained from the Y-intercept of the van't Hoff plot. ΔG^0 at a particular temperature T in Kelvin can be calculated from

$$\Delta G^0 = -RT \ln K_T = \Delta H^0 - T\Delta S^0 \quad 5$$

where R is the universal gas constant which is equal to 1.986 cal/mol K.

The concentration method

The concentration method utilizes a plot of T_m^{-1} versus ln c_{ts}, where T_m is the thermal melting temperature of the duplex and c_{ts} is the total strand concentration of PNA or DNA.

Since T_m is defined by the temperature where $\alpha = 0.5$ for a two state transition, combining Equations (3) and (4) yields:

$$\frac{1}{T_m} = \frac{R}{\Delta H^0}\ln c_{ts} + \frac{\Delta S^0 - R\ln 4}{\Delta H^0}. \quad 6$$

Thus, the thermodynamic parameters can be extracted from a linear fit to a plot of T_m^{-1} versus ln c_{ts} according to Equation (6) (21), where ΔH^0 is obtained from the slope of the linear fit and ΔS^0 from the Y-intercept.

The values of the thermodynamic parameters calculated by this method are thus independent of strand concentration, which is not the case with the hyperchromicity method described above.

RESULTS

PNA and DNA duplexes

In order to elucidate the properties of PNA duplexes as compared to those of iso-sequential DNA duplexes at reduced water activity, we have studied the effect of organic co-solvents on the thermal stability and thermodynamics of PNA·PNA, PNA·DNA and DNA·DNA duplexes of 'random', mixed base-sequence (Table S1). We chose DMF (dielectric constant of DMF is 36.7) and dioxane (dielectric constant of dioxane is 2.2) as organic co-solvents, as these are aprotic and are not hydrogen bond donors but still sufficiently polar to retain solubility of the PNA·DNA complexes even above 50% organic solvent.

The thermal stability of these duplexes and the corresponding thermodynamic parameters [evaluated using both the concentration method (T_m^{-1} versus $\ln c_t$ plot) (21) and the hyperchromicity (curve fitting) method (21)] in aqueous medium with increasing amount of DMF (extrapolated to 100% DMF) are presented in Tables 1 and S2. Representative thermal melting curves are shown in Figures S1 and S2. It is important to note that thermal denaturation curves showed essentially unperturbed monophasic behaviour up to 50% organic co-solvent. The extrapolation was performed on the basis of the linear plots of the thermal stability (T_m) and Gibbs' free energy changes (ΔG^0) as a function of increasing amount of DMF in the medium (Figure 1A and B). Because of insufficient thermal stability of the DNA1·DNA2 duplex in >30% DMF, we also designed a longer DNA·DNA duplex (DNA3·DNA4), that has thermal stability in aqueous solution similar to that of the PNA1·PNA2 duplex (data in Tables 1 and S3, Figure 1A and B, representative thermal curves in Figure S3). While the organic solvent-dependent changes in ΔG^0 of all these duplexes show a very good correlation with the corresponding T_m, ΔH^0 and ΔS^0 values were not significantly affected within experimental error (Tables S2–S4). The extraction of thermodynamic parameters derived from thermal melting method requires that no change in heat capacity occurs in the duplex single-strand equilibrium. For the present systems we have found only

Table 1. Thermal stability and thermodynamic parameters of PNA and DNA duplexes[a]

Duplex	DMF[b] (%)	T_m (°C)[c,d]	ΔG^0_{37} (kcal/mol)[d,e]	$\Delta\Delta G^{0\,f}$
PNA1·PNA2	0	70.2 ± 0.3	−16.6 ± 0.5 (−17.2)	5.6
	10	68.3 ± 0.2	−15.4 ± 0.5 (−16.2)	
	20	66.2 ± 0.2	−14.3 ± 0.4 (−15.6)	
	30	64.3 ± 0.5	−14.5 ± 0.8 (−15.7)	
	40	63.4 ± 0.3	−14.4 ± 0.5 (−15.9)	
	50	61.3 ± 0.4	−13.6 ± 0.7 (−15.5)	
	60	60.0 ± 0.6[g]	[h]	
	70	56.1 ± 0.5[g]	[h]	
	100[i]	52.6	−11.0	
DNA1·DNA2	0	35.8 ± 0.5	−8.2 ± 0.7 (−7.9)	15.0
	10	29.2 ± 0.4	−6.2 ± 1.2 (−6.1)	
	20	23.7 ± 0.5	−4.9 ± 0.8 (−4.9)	
	30[j]	18.2 ± 0.3	−3.7 ± 0.5 (−3.9)	
	100[i]	~22.9	6.8	
DNA3·DNA4	0	70.0 ± 0.4	−16.4 ± 0.4	19.2
	10	62.1 ± 0.5	−14.3 ± 0.7	
	20	56.2 ± 0.6	−12.8 ± 0.8	
	30	49.2 ± 0.3	−11.3 ± 1.0	
	40	41.3 ± 0.2	−9.3 ± 0.6	
	50	30.1 ± 0.5	−6.1 ± 0.5	
	100[i]	~6.2	2.8	
PNA1·DNA2	0	51.3 ± 0.4	−10.3 ± 1.0	6.8
	10	47.0 ± 0.5	−9.5 ± 0.7	
	20	43.1 ± 0.3	−9.0 ± 0.6	
	30	40.2 ± 0.4	−8.3 ± 0.6	
	40	36.2 ± 0.5	−7.7 ± 0.8	
	50	31.1 ± 0.2	−6.7 ± 0.5	
	100[i]	12.3	−3.5	

[a]PNA and DNA sequences: H-GTA GAT CAC T-Lys-NH_2 (PNA1); H-AGT GAT CTA C-Lys-NH_2 (PNA2); 5′-GTA GAT CAC T-3′ (DNA1); 5′-AGT GAT CTA C-3′ (DNA2); 5′-AGT GAT CTA CGG TGG ACG GTC C-3′ (DNA3); 5′-GGA CCG TCC ACC GTA GAT CAC T-3′ (DNA4).
[b]Vol% in 10 mM phosphate buffer containing 100 mM NaCl and 0.1 mM EDTA, pH 7.2 ± 0.01.
[c]Duplex concentrations of 5.0 μM in strands were used (T_m plots are in Figures 1A, S4A and S5A).
[d]Standard deviations are based on five independent measurements.
[e]Evaluated from the hyperchromicity (curve fitting) method (21) at 37°C (ΔG^0_{37} plots are in Figures 1B, S4B and S5B). Values obtained from the concentration method (21) (Equation 6) are shown in parantheses (full details of these data are in Supplementary Data).
[f]$\Delta\Delta G^0 = \Delta G^0_{37}$ (DMF) − ΔG^0_{37} (aqueous). Calculated only with the values obtained from the hyperchromicity method.
[g]These data have poorer accuracy due to upper baseline irregularities.
[h]Data could not be evaluated because of bad thermal curves.
[i]Data were obtained from manual extrapolation of the linear plots in Figure 1A and B.
[j]Data obtained at higher than 30% of DMF were not reliable because of too low values of T_m.

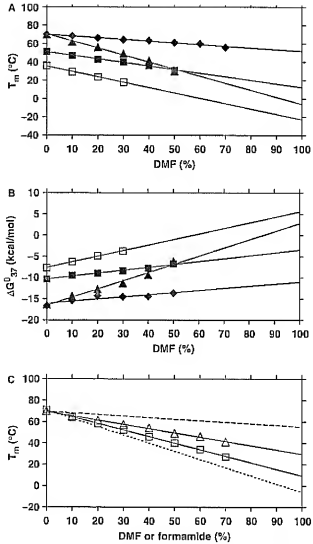

Figure 1. Plots of (A) T_m and (B) ΔG^0 of PNA1·PNA2 (*solid diamond*), PNA1·DNA2 (*solid square*), DNA1·DNA2 (*open square*) and DNA3·DNA4 (*solid triangle*) as a function of the amount of DMF in the medium. (C) Plots of T_m of PNA1·PNA2 (*open triangle*) and DNA3·DNA4 (*open square*), as a function of the amount of formamide in the medium, compared to that of PNA1·PNA2 (dashed line) and DNA3·DNA4 (dotted line), as a function of the amount of DMF in the medium (DMF data taken from Figure 1A). The aqueous buffer was 10 mM phosphate buffer containing 100 mM NaCl and 0.1 mM EDTA, pH 7.2 ± 0.01 (data in Tables 1, S2–S5).

minor changes. The values of specific heat capacity change (enthalpic) at constant pressure (ΔC_p) were 2.0 kcal/mol.K for PNA1·PNA2, 1.4 kcal/mol.K for PNA1·DNA2, 1.9 kcal/mol.K for PNA2·DNA1 and 1.8 kcal/mol.K for DNA1·DNA2 (19).

It is clearly evident that the presence of DMF has a much smaller effect on the thermal stability and free energy of the PNA·PNA duplex than on the DNA·DNA duplex, whereas the effect on the PNA·DNA duplex is intermediate. The plots in Figure 1A and B show a linear dependence of both T_m and ΔG^0 on the DMF concentration. The change in thermal stabilities as a function of the amount of DMF in the medium did not deviate within experimental error from linearity up to 70% of DMF, thereby supporting a hypothetical approximation beyond 70% of DMF (and an extrapolation of the data to 100% DMF) (Figure 1A and B, Table 1). These results strongly indicate that PNA·PNA duplexes in contrast to DNA·DNA (and PNA·DNA) duplexes will have appreciable stability even in the absence of water. Naturally, we cannot exclude that non-linearity might occur at very low water contents.

Analogous studies of PNA1·PNA2, PNA1·DNA2 and DNA3·DNA4 in dioxane show that, despite the significant difference between the dielectric constants of these solvents, there is no significant difference between the effects of DMF and dioxane on T_m and ΔG^0 (Tables S3 and S4, representative thermal curves in Figures S2 and S3, plots in Figures S4 and S5).

PNA and DNA duplexes in formamide

In order to compare the effect of aprotic solvent (DMF and dioxane, where the destabilization of the duplex is assumed to be predominantly caused by dehydration (and perhaps change in dielectrics) with that of a hydrogen-bond donor (and breaking) solvent, we studied the thermal and thermodynamic properties of PNA1·PNA2 and DNA3·DNA4 in formamide (dielectric constant of formamide is 109), which is a well-established nucleic acid denaturant (Table S5, representative thermal curves in Figure S6, plots in Figures 1C, S7 and S8). In formamide, the destabilization is caused by a combined effect of H-bond disruption and dehydration. Notably, the destabilizing effect of formamide is almost as pronounced for the PNA duplex as for the DNA duplex. The slopes of the linear plots of T_m of PNA1·PNA2 and DNA3·DNA4 as a function of increasing amount of DMF in the medium are −0.17 (without taking the values at 60 and 70% into account) and −0.77, respectively, whereas, those values in formamide are −0.40 and −0.61, respectively (Figure 1C). On the other hand, the decrease in T_m is paralleled by the increases in both ΔG^0 and ΔH^0. Thus in stark contrast to the effects of DMF and dioxane, formamide has comparable effects on PNA and DNA duplex stabilities.

Self-complementary PNA and DNA hairpins

Additionally, we examined the effect of organic co-solvent on the stability of self-complementary (foldback) PNA hairpins (Figure 2). Unfolding of hairpin PNAs, H-AGAG-(eg1)$_3$-CTCT-Lys-NH$_2$ (PNA3) and H-ACAG-(eg1)$_3$-CTGT-Lys-NH$_2$ (PNA4) also showed little change in T_m and ΔG^0 with increasing amount of DMF or dioxane in the medium (Tables S6 and S7, Figures 2A, B and S9). In fact, the hairpins were slightly stabilized in dioxane (Table S7 and Figure 2B), and in this case the stabilization appears to be enthalpic (Table S7). Similar studies with an analogous DNA hairpin (DNA5: 5′-AGAGTTTTCTCT-3′) in DMF and dioxane as a control showed a dramatic thermal destabilization by both co-solvents (Tables S8 and S9, Figures 2C, D and S10).

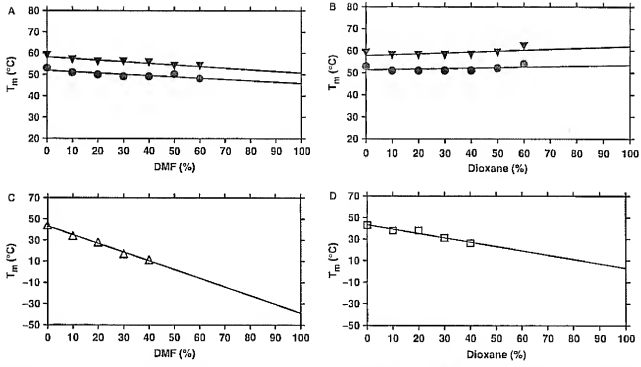

Figure 2. Plots of T_m of self-complementary (hairpin) PNAs PNA3 (*solid inverted triangle*) and PNA4 (*solid circle*) as a function of the amount of (A) DMF and (B) dioxane in the medium. Plots of T_m of hairpin DNA control DNA5 as a function of the amount of (C) DMF (*open triangle*) and (D) dioxane (*open square*) in the medium. The aqueous buffer was 10 mM phosphate buffer containing 100 mM NaCl and 0.1 mM EDTA, pH 7.2 ± 0.01 (PNA3: H-AGAG-(eg1)$_3$-CTCT-Lys-NH$_2$, PNA4: H-ACAG-(eg1)$_3$-CTGT-Lys-NH$_2$, DNA5: 5′-AGA GTT TTC TCT-3′) (data in Tables S6–S9).

PNA duplex with an endstacked tricyclic thymine

In order to evaluate the effect of removal of water from the medium (by increasing DMF content) on base stacking forces more directly, we studied the effect of DMF on the behaviour of a PNA duplex (PNA5·PNA6) containing a tricyclic thymine analogue (benzo[*b*]-1,8-naphthyridin-2(1*H*)-one, designated as tT) (22) attached at the end of one strand, thereby stabilizing the duplex predominantly (or entirely) by end-stacking (22). It is evident that, DMF preferentially destabilizes the PNA5·PNA6 duplex (T_m slope is −0.24) compared to the control PNA6·PNA7 duplex (T_m slope is −0.15) (Table S10, Figures 3 and S11). Most interestingly, the stabilization of the duplex contributed by the tT base is completely lost at 70% DMF (Figure 3), clearly indicating a dramatic reduction of the contribution of stacking interactions to duplex stability under these conditions (decrease in thermal stability of PNA5·PNA6 at 70% DMF with respect to 0% DMF, $\Delta T_m = 18.8$, whereas, that of the control duplex PNA6·PNA7 is only 10.7).

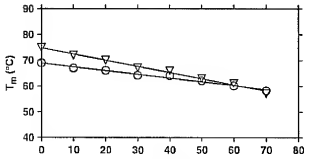

Figure 3. Plots of T_m of PNA5·PNA6 (*open inverted triangle*) containing tricyclic thymine (tT) and its control PNA6·PNA7 (*open circle*), as a function of the amount of DMF in the medium (data in Table S10). The aqueous buffer was 10 mM phosphate buffer containing 100 mM NaCl and 0.1 mM EDTA, pH 7.2 ± 0.01 (PNA5: H-tT-GTA GAT CAC T-NH$_2$, PNA6: H-AGT GAT CTA C-NH$_2$, PNA7: H-GTA GAT CAC T-NH$_2$).

Mismatched duplexes

Sequence discrimination is critically dependent on hydrogen bonding recognition, but base-pair mismatches may also change the geometry of the helix and thus influence stacking interactions. For instance, X-ray crystallography and NMR studies have revealed that G·T and T·T mismatches in DNA likely form wobble base pairs (23–30). The homopurine A·A mismatch in DNA involves one hydrogen bond between the amino group of one adenine and the nitrogen N1 of the opposite adenine, whereas, the mispaired T·T bases in DNA involve two imino-carbonyl hydrogen bonds (23–30). A C·T mismatch is the most unstable in DNA and makes one regular

H-bond involving carbonyl groups along with one weak H-bond bridged by a water molecule at neutral pH (23–30). Effects of mispairs are rather localized and the duplex retains a global B-form conformation (23–30). As hydration and stacking forces could be more important factors for mismatch stabilization, removal of water may preferentially destabilize these.

However, surprisingly we find that the relative effects of DMF on the stability of single base mismatched (T·T, A·A and C·T) duplexes of PNA8 (H-AGT GTT CTA C-Lys-NH_2), PNA9 (H-GTA GAA CAC T-Lys-NH_2) and PNA10 (H-GTA GCT CAC T-Lys-NH_2) with PNA1 and PNA2 are not distinguishable within experimental error from that observed with the corresponding fully matched PNA duplexes (The T_m slope of the fully matched PNA1·PNA2 is −0.17, of T·T mismatched PNA1·PNA8 is −0.13, of A·A mismatched PNA2·PNA9 is −0.08, and of C·T mismatched PNA2·PNA10 is −0.12) (Tables 1 and S11, Figures 1A, 4, S12 and S13). These results suggest that, in these base pair mismatched duplexes, hydrogen bonding and stacking interactions are equally compromised or that the mismatched base pair contributes only very little to the overall duplex stability.

DISCUSSIONS AND CONCLUSIONS

The relative insensitivity of PNA duplexes to the reduced water content is consistent with our previous observations showing that in contrast to the case of DNA·DNA (and to a lesser extent PNA·DNA duplexes), no change in hydration occurs upon duplex formation (19) and no change in counter ion binding is observed either (19).

Thus the effect of organic solvents on PNA duplex stability may be ascribed to changes in base-pair hydrogen bonding and stacking forces of the base pairs. Stacking forces have contributions from hydrophobic, dispersion and dipole electrostatic forces (31–33), and therefore both hydrogen bonding as well as stacking interactions are affected by the dielectric constant of the solvent. However, we observe no correlation between PNA duplex stability and solvent dielectric constant. Water, formamide, DMF and dioxane have dielectric constants of 78, 109, 37 and 2.2, respectively, and in particular the comparable effect of dioxane and DMF strongly argue against a simple dependence of the dielectric constant. On the other hand, it is clear that the contribution of hydrogen bonding to PNA duplex stability should significantly increase when the dielectric constant and especially the hydrogen bonding donor/acceptor activity of the solvent decreases.

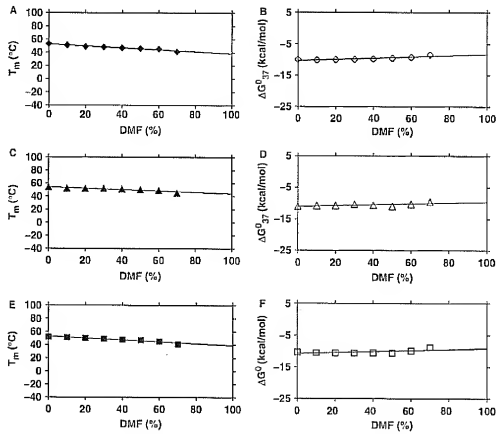

Figure 4. Thermal stabilities of T·T, A·A and C·T mismatched PNA duplexes in DMF (data in Table S11). Plots of (A) T_m (*solid diamond*) and (B) ΔG^0 (*open diamond*) of PNA1·PNA8 as a function of the amount of DMF in the medium. Plots of (C) T_m (*solid triangle*) and (D) ΔG^0 (*open triangle*) of PNA2·PNA9 as a function of the amount of DMF in the medium. Plots of (E) T_m (*solid square*) and (F) ΔG^0 (*open square*) of PNA2·PNA10 as a function of the amount of DMF in the medium. The aqueous buffer was 10 mM phosphate buffer containing 100 mM NaCl and 0.1 mM EDTA, pH 7.2 ± 0.01.

(It should be kept in mind that the hydrogen bond stabilization results as the difference between nucleobase hydrogen bonding to solvent versus base pairing.) Furthermore, our results very clearly show that the stabilizing effect of end stacking decreases upon addition of DMF to the solvent. Indeed, the end-stacking effect using a tricyclic thymine analogue is absent already at 70% DMF. This thymine analogue in essence contains an additional phenyl ring compared to thymine (22), and therefore the stacking interactions of the tricyclic thymine are expected to contain a relatively larger hydrophobic contribution than natural nucleobase pairs (32). Thus a similar effect on the natural nucleobases is expected to occur at higher DMF (lower water) content. Nonetheless, our results strongly support the contention that relative helix stabilizing contribution of base-pair hydrogen bonding and base stacking will be shifting towards hydrogen bonding as the water content of the medium (as well as the dielectric constant) decreases. In this context, we note that for the PNA hairpins an increased stability is indicated with increasing concentration of the low dielectric constant solvent dioxane (in contrast to DMF).

It is well established that formamide and urea destabilize DNA and RNA duplexes (34,35), and in agreement with the present data, the effect of DMF and dioxane (at 10 mol%) on RNA duplex stability was found to be comparable to that of formamide (33). However, also in case of RNA, no direct correlation between the duplex stabilizing effect and the solvent dielectric constant was apparent (33). Indeed, from our data comparing the effects of formamide and DMF on PNA duplex stability, we would argue that the effects of organic solvents on DNA and RNA duplexes is to a larger extent due to hydration (solvation) and counterion effects, whereas the effects on PNA duplexes more directly reflect changes in hydrogen bonding and stacking contributions to helix stability. Thus more systematic studies on the different behaviour of PNA versus DNA duplexes could possibly shed light on these relative contributions to nucleobase (including DNA) helix stability.

Interestingly, the data (Tables S2 and S3) could hint that the decreased stability of the DNA duplexes upon removal of water appears to be primarily entropic rather than enthalpic, though the data are not conclusive. This would be compatible with the differences in behaviour between the PNA and the DNA duplexes (as mentioned above) being caused by the differences in hydration and counter ion release rather than via nucleobase interaction. Thus, these results can be of great significance in understanding nucleobase pairing systems, and it will be of great interest to learn whether other non-ionic DNA analogues such as methylphosphonates (8), phosphotriester (9) and morpholino (10) derivatives behave similarly.

The possibility of having stable nucleobase paired double helices in organic solvents has several implications. First of all, we note that it should be possible to reach the goal of finding PNA duplexes that are freely soluble in non-polar organic solvents by chemically modifying the PNA backbone, for instance using hydrophobic amino acids in place of glycine (36). Studying the behaviour of these in the absence of water could provide valuable information about fundamental properties of nucleobase double helices. Furthermore, PNA duplexes spanning lipid membranes could be envisaged as a means for mediating electron transport over the membrane. Sequence-dependent electron transport (or hole migration) through DNA duplexes is now well established (37–41), and although electron transport involving PNA still require more extensive studies (42,43) this should be worthwhile pursuing. Finally, entirely new dimensions would be added to the emerging technique of 'DNA sequence' directed organic synthesis (44). Thus novel avenues of exploiting nucleobase recognition systems can be enabled by transferring these to non-aqueous environments.

SUPPLEMENTARY DATA

Supplementary data are available at NAR Online.

ACKNOWLEDGEMENT

This work is supported by the European Commission in the 6th Framework PACE project, contract no. 002035. Funding to pay the Open Access publication charges for this article was provided by the European Commission.

Conflict of interest statement. None declared.

REFERENCES

1. Watson,J.D. and Crick,F.H. (1953) Genetical implications of the structure of deoxyribonucleic acid. *Nature*, 171, 964–967.
2. Saenger,W. (1988) *Principles of Nucleic acid Structure*. Springer, New York.
3. Kool,E.T. (2001) Hydrogen bonding, base stacking, and steric effects in DNA replication. *Annu. Rev. Biophys. Biomol. Struct.*, 30, 1–22.
4. Ostenbrink,C. and van Gunsteren,W.F. (2005) Efficient calculation of many stacking and pairing free energies in DNA from a few molecular dynamics simulations. *Chemistry*, 11, 4340–4348.
5. Sundaralingam,M. and Ponnuswamy,P.K. (2004) Stability of DNA duplexes with Watson-Crick base pairs: a predicted model. *Biochemistry*, 43, 16467–16476.
6. Protozanove,E., Yakovchuk,P. and Frank-Kamenetskii,M.D. (2004) Stacked-unstacked equilibrium at the nick site of DNA. *J. Mol. Biol.*, 342, 775–785.
7. Yakovchuk,P., Protozanove,E. and Frank-Kamenetskii,M.D. (2006) Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. *Nucleic. Acids Res.*, 34, 564–574.
8. Miller,P.S, Yano,J., Yano,E., Carroll,C., Jayaraman,K. and Ts'o,P.O. (1979) Nonionic nucleic acid analogues. Synthesis and characterization of dideoxyribonucleoside methylphosphonates. *Biochemistry*, 18, 5134–5143.
9. Tosquellas,G., Alvarez,K., Dell'Aquila,C., Morvan,F., Vasseur,J.-J., Imbach,J.-L. and Rayner,B. (1998) The pro-oligonucleotide approach: solid phase synthesis and preliminary evaluation of model pro-dodecathymidylates. *Nucleic Acids Res.*, 26, 2069–2074.
10. Summerton,J. (1999) Morpholino antisense oligomers: the case for an RNase H-independent structural type. *Biochim. Biophys. Acta*, 1489, 141–158.
11. Nielsen,P.E., Egholm,M., Berg,R.H. and Buchardt,O. (1991) Sequence-selective recognition of DNA by strand displacement with a thymine-substituted polyamide. *Science*, 254, 1497–1500.
12. Nielsen,P.E. (1999) Peptide nucleic acid. A molecule with two identities. *Acc. Chem. Res.*, 32, 624–630.

13. Egholm,M., Buchardt,O., Christensen,L., Behrens,C., Freier,S.M., Driver,D.A. *et al.* (1993) PNA hybridizes to complementary oligonucleotides obeying the Watson-Crick hydrogen-bonding rules. *Nature*, 365, 566–568.
14. Wittung,P., Nielsen,P.E., Buchardt,O., Egholm,M. and Nordén,B. (1994) DNA-like double helix formed by peptide nucleic acid. *Nature*, 368, 561–563.
15. Jensen,K.K., Orum,H., Nielsen,P.E. and Nordén,B. (1997) Kinetics for hybridization of peptide nucleic acids (PNA) with DNA and RNA studied with the BIAcore technique. *Biochemistry*, 36, 5072–5077.
16. Brown,S.C., Thomson,S.A., Veal,J.M. and Davis,D.G. (1994) NMR solution structure of a peptide nucleic acid complexed with RNA. *Science*, 265, 777–780.
17. Eriksson,M. and Nielsen,P.E. (1996) Solution structure of a peptide nucleic acid-DNA duplex. *Nature Struct. Biol.*, 3, 410–413.
18. Rasmussen,H., Kastrup,J.S., Nielsen,J.N., Nielsen,J.M. and Nielsen,P.E. (1997) Crystal structure of a peptide nucleic acid (PNA) duplex at 1.7 A resolution. *Nature Struct. Biol.*, 4, 98–101.
19. Sen,A. and Nielsen,P.E. (2006) Unique properties of purine/pyrimidine asymmetric PNA.DNA duplexes: differential stabilization of PNA.DNA duplexes by purines in the PNA strand. *Biophys. J.*, 90, 1329–1337.
20. Christensen,L., Fitzpatrick,R., Gildea,B., Petersen,K., Hansen,H.F., Koch,T., Egholm,M., Buchardt,O., Nielsen,P.E. *et al.* (1995) Solid-phase synthesis of peptide nucleic acids. *J. Pept. Sci.*, 1, 175–183.
21. Peyret,N., Seneviratne,P.A., Allawi,H.T. and SantaLucia,J.Jr (1999) Nearest-neighbor thermodynamics and NMR of DNA sequences with internal A.A, C.C, G.G, and T.T mismatches. *Biochemistry*, 38, 3468–3477.
22. Eldrup,A., Nielsen,B.B., Hnatowich,G., Rasmussen,H., Kastrup,J.S., Christensen,C. and Nielsen,P.E. (2001) 1,8-Naphthyridin-2(1*H*)-ones—novel bicyclic and tricyclic analogues of thymine in peptide nucleic acids (PNAs). *Eur. J. Org. Chem.*, 9, 1781–1790.
23. Patel,D.J., Shapiro,L. and Hare,D. (1987). In Eckstein,F. and Lilley,D.M.J. (eds), *In Nucleic acid and Molecular Biology*. Springer, Berlin, pp. 70–84.
24. Germann,M.W., Kalisch,B.W., Varnum,J.M., Vogel,H.J. and van de Sande,J.H. (1988) NMR spectroscopic and enzymatic studies of DNA hairpins containing mismatches in the EcoRI recognition site. *Biochem. Cell Biol.*, 76, 391–402.
25. Allawi,H.T. and SantaLucia,J.Jr (1998) Thermodynamics of internal C.T mismatches in DNA. *Nucleic. Acids Res.*, 26, 2694–2701.
26. Reulard,Y., Cognet,J.A.H. and Fazakerley,G.V. (1997) Solution structure as a function of pH of two central mismatches, C.T and C.C, in the 29 to 39 K-ras gene sequence, by nuclear magnetic resonance and molecular dynamics. *J. Mol. Biol.*, 268, 331–347.
27. Nemecek,D., Stepanek,J., Turpin,P.Y. and Rosenberg,I. (2004) Raman study of potential 'antisense' drugs: nonamer oligonucleotide duplexes with a central mismatch as a model system for the binding selectivity evaluation. *Biopolymers*, 74, 115–119.
28. Hunter,W.N., Brown,T., Kneale,G., Anand,N.N., Rabinovich,D. and Kennard,O. (1987) The structure of guanosine-thymidine mismatches in B-DNA at 2.5-A resolution. *J. Biol. Chem.*, 262, 9962–9970.
29. Ke,S.-H. and Wartell,R.M. (1996) The thermal stability of DNA fragments with tandem mismatches at a d(CXYG).d(CY'X'G) site. *Nucleic. Acids Res.*, 24, 707–712.
30. Gervais,V., Cognet,J.A., Le Bret,M.L., Sowers,C. and Fazukerley,G.V. (1995) Solution structure of two mismatches A.A and T.T in the K-ras gene context by nuclear magnetic resonance and molecular dynamics. *Eur. J. Biochem.*, 228, 279–290.
31. Guckian,K.M., Schweitzer,B.A., Ren,R.X.-F., Sheils,C.J., Tahmassebi,D.C. and Kool,E.T. (2000) Factors contributing to aromatic stacking in water: Evaluation in the context of DNA. *J. Am. Chem. Soc.*, 122, 2213–2222.
32. Kim,T.W. and Kool,E.T. (2005) A series of nonpolar thymidine analogs of increasing size: DNA base pairing and stacking properties. *J. Org. Chem.*, 70, 2048–2053.
33. Hickey,D.R. and Turner,D.H. (1985) Solvent effects on the stability of A_4U_4p. *Biochemistry*, 24, 2086–2094.
34. Hutton,J.R. (1977) Renaturation kinetics and thermal stability of DNA in aqueous solutions of formamide and urea. *Nucleic. Acids Res.*, 4, 3537–3555.
35. Blake,R.D. and Delcourt,S.G. (1996) Thermodynamic effects of formamide on DNA stability. *Nucleic. Acids Res.*, 24, 2095–2103.
36. Pfleiderer,W., Sforza,S., Haulina,G., Dahl,O. and Nielsen,P.E. (1998) Peptide nucleic acids (PNA) with a functional backbone. *Tetrahedron Lett.*, 39, 4707–4710.
37. Schuster,G.B. (2004) *In Long-Range Charge Transfer in DNA I-II*. Springer, Heidelberg.
38. Delaney,S. and Barton,J.K. (2003) Long-range DNA charge transport. *J. Org. Chem.*, 68, 6475–6483.
39. Fink,H.W. and Schonenberger,C. (1999) Electrical conduction through DNA molecules. *Nature*, 398, 407–410.
40. Wan,C., Fiebig,T., Schiemann,O., Barton,J.K. and Zewail,A.H. (2000) Femtosecond direct observation of charge transfer between bases in DNA. *Proc. Natl Acad. Sci. USA*, 97, 14052–14055.
41. Shao,F., Augustyn,K. and Barton,J.K. (2005) Sequence dependence of charge transport through DNA domains. *J. Am. Chem. Soc.*, 127, 17445–17452.
42. Armitage,B., Ly,D., Koch,T., Frydenlund,H., Orum,H., Batz,H.G. and Schuster,G.B. (1997) Peptide nucleic acid-DNA duplexes: long range hole migration from an internally linked anthraquinone. *Proc. Natl Acad. Sci. USA*, 94, 12320–12325.
43. Tanabe,K., Yoshida,K., Dohno,C., Okamoto,A. and Saito,I. (2000) Control of electron transfer in DNA by peptide nucleic acids (PNA). *Nucleic Acids Symp. Ser.*, 44, 35–36.
44. Gartner,Z.J., Tse,B.N., Grubina,R., Doyon,J.B., Snyder,T.M. and Liu,D.R. (2004) DNA-templated organic synthesis and selection of a library of macrocycles. *Science*, 305, 1601–1605.